South Carolina
Alternate Assessment
(SC-Alt)

South Carolina's Alternate Assessment, SC-Alt
Spring 2008 Operational Administration

Technical Report
Unedited Draft

American Institutes for Research
South Carolina Department of Education

Table of Contents

Index of Exhibits

Introduction

This report details the design, development, and spring 2008 operational test results for the South Carolina Alternate Assessment (SC-Alt). The SC-Alt assessment consists of four content areas: English language arts (ELA), mathematics, science, and social studies. The assessments are administered across three grade-bands: 3–5, 6–8, and 10.

The first chapter of this Technical Report describes the background of the alternate assessments in South Carolina, the format of the previous assessments, and the need for a new alternate assessment.

The second chapter is comprehensive in its scope and includes information on the design of the alternate assessment and the development of tasks and items to measure academic growth among students with significant cognitive disabilities. The field-test designs are also summarized in the second chapter.

Chapter 2 further reviews how the design of the alternate assessment is unique in that it uses a Student Placement Questionnaire (SPQ) to maximize the efficiency of teacher and student testing time by guiding the teacher to administer tasks at a complexity level appropriate for the achievement level of each individual student. A thorough review of the SPQ is presented in Chapter 2.

Chapter 2 also reviews how the design of the assessment allows for the development of a vertical scale by linking grade-appropriate tasks across grade levels and complexity levels within grades. A vertical scale presents many benefits to the assessment system; these benefits will be discussed later in this report in the sections on scaling and score reporting.

Chapter 3 details the spring 2008 operational test administration in ELA, mathematics, science, and social studies; test administrator training; use of the SPQ; measures taken to ensure the accuracy of scoring; and the maintenance of test security.

Chapter 4 describes the standard-setting procedures conducted to establish performance standards. The chapter includes a description of the Item-Descriptor (ID) Matching procedure, the goals of the standard-setting workshop, the composition of the standard-setting panels, the workshop activities, and the panels' recommended performance standards.

Chapter 5 reviews several technical topics, including analysis and scaling and the reliability of test scores. This chapter includes a description of the procedures used to calculate internal consistency reliability estimates and classification accuracy estimates.

Chapter 6 describes the score reporting system for SC-Alt and references the Individual Student (Family) Report (included in Appendix H) from which the summary reports are derived. The chapter provides a brief description of the score reports, their intended uses, and the information they contain.

Chapter 7 provides an overview of statewide achievement on the SC-Alt, based on the spring 2008 operational test administration.

Chapter 8 reports on content validity and convergent and discriminant validity topics as well as the validity of the SPQ.

Chapter 1: History of the Development of Alternate Assessment in South Carolina

Overview of the State Assessment System

The South Carolina Assessment System includes the South Carolina Palmetto Achievement Challenge Tests (PACT), the High School Assessment Program (HSAP), the End-of-Course Examination Program (EOCEP), and the South Carolina Readiness Assessment (SCRA). These state-level assessments are required by the Education Accountability Act of 1998 (EAA) and are aligned with the state's academic standards for each subject and grade level.

- PACT measures the performance of all public school students in grades 3 through 8 in the content areas of English language arts (ELA), mathematics, science, and social studies.
- HSAP measures the performance of high school students in ELA and mathematics and is used both as one criterion for eligibility to receive a high school diploma and as the primary source for reporting the federally mandated data required by the No Child Left Behind Act (NCLB).
- EOCEP is administered in gateway courses at the high school level. The physical science EOCEP examination is counted for participation purposes for NCLB reporting.
- SCRA, an assessment of student readiness, is administered to students in kindergarten and first grade. This is a teacher rating scale, and the results are not included in the state accountability system.

The EAA establishes a performance-based accountability system that includes all students. This act supports South Carolina's commitment to public education and a conviction that high expectations for all students are a vital component of improving academic education.

The goals of the state assessment system are as follows:

- ***increasing academic performance of all children and, ultimately, raising high school graduation rates;***
- ***implementing rigorous academic achievement standards that are aligned with the South Carolina curriculum standards;***
- ***improving instruction based, in part, on the implementation of these higher standards; and***
- ***using the results of challenging assessments that measure student performance relative to these standards.***

Another goal is to inform various audiences—teachers, school administrators, district administrators, South Carolina State Department of Education (SCDE) staff, parents, and the public—of the status of academic performance and of the progress of public school students toward meeting South Carolina's academic achievement standards.

The South Carolina academic standards form the basis for alignment across the state education system for district and school curricula, classroom instruction, units of study, and learning experiences. The academic standards are the basis for all assessments in the state assessment system, including alternate assessment.

Overview and Purpose of the South Carolina Alternate Assessment

The South Carolina alternate assessment system received Full Approval with Recommendations from the U.S. Department of Education (USDOE) under Title I of the Elementary and Secondary Education Act (ESEA), as amended by the No Child Left Behind Act of 2001 (NCLB) in February 2006, following Peer Review. The letter of approval stated that the alternate assessment met essential regulatory requirements but recommended strengthening the alternate assessment. SCDE had already begun to address revisions to the alternate assessment system on the basis of experience with the existing system and a better understanding of the requirement to align instruction and assessment to grade-level academic standards.

The purpose of the alternate assessment based on alternate achievement standards is to capture and evaluate the performance of students who have traditionally been excluded from statewide testing programs and to improve instruction for these students by promoting appropriately high expectations and the inclusion of these students in state accountability for district report cards and for Adequate Yearly Progress (AYP) reporting at the school, district, and state levels.

Description of the South Carolina Alternate Assessment

The SC-Alt is administered to students who have been determined by the Individualized Education Program (IEP) team to be unable to participate in the general state assessments even with appropriate accommodations. It is an alternate assessment on alternate achievement standards to the PACT for students in grades 3–8 and the HSAP and Physical Science EOCEP for high school students. An alternate scoring format is provided for the SCRA for students who are the ages of typical students in kindergarten and first grade. Information regarding the alternate scoring may be found in the SCRA documentation.

The test is administered to students who meet the participation criteria for alternate assessment and who are of the ages of typical students in grades 3–8 and 10. Students who are ages 8–13 (the typical ages for grades 3–8) are assessed in ELA, mathematics, science, and social studies. Students who are age 15 (the typical age of students in grade 10) are assessed in ELA, mathematics, and physical science.

The SC-Alt consists of a series of performance tasks that are scored by the test administrator (teacher) as they are administered. The performance tasks are scripted activities, and each task contains four to eight related items. The items have a scaffolded scoring script to reduce the complexity of the item when students do not respond successfully on the first attempt. All items are linked to the South Carolina academic content standards through the South Carolina Alternate Assessment Standards and Measurement Guidelines (ASMGs). The ASMGs are linked explicitly to the South Carolina academic standards for grades 3–8 and 10, although at less-complex or prerequisite levels.

The SC-Alt has three forms: elementary, middle, and high school. Students are assigned to forms on the basis of their age on September 1 of the tested year. Students who are ages 8–10 are assigned to the elementary school form, students who are ages 11–13 are assigned to the middle school form, and students who are age 15 participate in the high school form.

The assessment is designed to minimize teacher and student testing burden by administering only those items that are well-suited to a student's achievement level. The test administrator completes a Student Placement Questionnaire (SPQ) to determine the most appropriate starting task for the student. Tasks are arranged in order of difficulty (from easiest to most difficult). Once the appropriate starting task is identified, test administrators continue to administer tasks until the student can no longer respond successfully.

The first operational administration of the SC-Alt was conducted during a seven-week testing window during spring 2007 in ELA, mathematics, and science. A census field test was conducted during the same assessment window for the social studies assessment. Documentation related to the 2008 operational administration is the focus of this Technical Report.

Background on Alternate Assessment Development in South Carolina

The 1997 amendments to the Individuals with Disabilities Education Act (IDEA '97) created the mandate to include all children, including children with significant disabilities, in state testing and accountability systems. The vision for the South Carolina alternate assessment system was initiated in early 1998 in response to the IDEA '97 regulations. This vision has driven the development and revision of alternate assessment in South Carolina.

A core team of staff from the SCDE Offices of Exceptional Children, Assessment, Research, and Curriculum and Standards met in March 1998 to develop a plan for designing an alternate assessment to meet the IDEA mandate and to be included in the state assessment system. The team's first steps were to convene a steering committee and seek technical assistance from the Mid-South Regional Resource Center (MSRRC) to explore strategies for designing an alternate assessment.

The Alternate Assessment Steering Committee was convened on May 12, 1998, to assist SCDE in determining how to include students with significant cognitive disabilities in statewide assessments. The committee comprised parents, special education and general education teachers, administrators, and representatives from other agencies. Dr. Ken Olsen of MSRRC provided the committee with technical assistance, including information on IDEA requirements, examples of options that some states were using or considering, and research available on alternate assessment. He facilitated a process that allowed the Steering Committee to reach shared foundational beliefs, address eligibility criteria and content and performance standards, and develop plans.

To ensure that all students, including students with significant disabilities, are included in the testing and accountability systems and have appropriate access to instruction in the South Carolina academic standards, the Steering Committee determined that the alternate assessment would be based on the following principles:

- All children can learn, be expected to meet, and be challenged to meet high standards.
- Special education is an extension and adaptation of the general education program and curriculum, rather than an alternate or separate system.

- The South Carolina State Board-approved standards are the foundation for all students, including students with unique needs and abilities.
- Measurement and reporting must be defensible in terms of feasibility, validity, reliability, and comparability.
- Results of the state standards-based program must be used to improve planning, instruction, and learning.
- An alternate assessment is appropriate for the few students for whom the state assessment, even with accommodations, is not appropriate.
- The alternate assessment is designed for a diverse group of students and should be flexible enough to address their individual needs.

The committee articulated these goals for the alternate assessment:

- to provide evidence that students have acquired the skills and knowledge necessary to become as independent as possible;
- to document the student's performance and the performance of the programs serving the student;
- to merge instructional "best practice," instruction in state standards, and assessment activities; and
- to provide information in the development of curriculum that is responsive to the student's needs.

The Steering Committee created the following participation guidelines to guide IEP team decisions regarding students who should participate in the alternate assessment:

- The student demonstrates a significant cognitive disability and adaptive skills, which result in performance that is substantially below grade-level achievement expectations even with the use of accommodations and modifications.
- The student accesses the state-approved curriculum standards at less complex levels and with extensively modified instruction.
- The student has current adaptive skills requiring extensive direct instruction and practice in multiple settings to accomplish the application and transfer of skills necessary for application in school, work, home, and community environments.
- The student is unable to apply or use academic skills across natural settings when instructed solely or primarily through classroom instruction.
- The student's inability to achieve the state grade-level achievement expectations is not the result of excessive or extended absences or social, cultural, or economic differences.

NOTE: The term *significant cognitive disabilities* was added by the South Carolina Alternate Assessment Advisory Committee to the criteria after the passage of the NCLB December 2003 regulations on alternate assessment.

The Steering Committee recommended that the state develop a portfolio collection of evidence of student progress toward the South Carolina academic standards similar in design to the Kentucky Portfolio Alternate Assessment. The committee also recommended that SCDE

prepare a Request for Proposal (RFP) for a contractor to develop the alternate assessment. Advanced Systems in Measurement and Evaluation Inc. (ASME), which later became Measured Progress, was awarded the contract. This company, along with the Inclusive Large Scale Standards and Assessment (ILSSA) project at the University of Kentucky, began work with SCDE on the design of PACT-Alt.

A work group was convened to define the domain for instruction and assessment. To ensure that the South Carolina curriculum standards were the foundation for all students, including students with unique needs and abilities, the work group developed adaptations of the curriculum standards. The work group comprised special education teachers, regular education teachers, parents, administrators, higher education personnel, representatives from community agencies, and SCDE personnel. The work group process, which was facilitated by staff from MSRRC, focused on the prerequisite skills found primarily in the curriculum standards in prekindergarten through grade 2.

The work group affirmed that special education services must operate as an extension of the general education program and curriculum rather than as an alternate or separate system. The standards in this initial document were identified as concepts that every student, including students with moderate to severe disabilities, should know or be able to perform. These selected standards, which focused on skills that were deemed essential and attainable for every student, were directed toward the following goals:

- enhancing the quality of students' communication skills;
- improving the quality of students' everyday living;
- improving students' ability to function in society and promote in them an acceptance of and respect for self and others;
- preparing students for transition into adult living; and
- moving students toward independence, which may range from a level of self-care with assistance to total self-sufficiency.

The extensions were based on the state academic content standards in prekindergarten through grade 2. For each selected standard, examples of essential real-world performance skills were developed. The articulation of these performance skills was designed to provide the rationale for teaching the standards and to serve as guides for teachers and parents regarding what the skill "looked like" when a student demonstrated it. The committee specified that these performance skills could be accomplished in home, school, and community environments through a variety of individualized communication systems and might incorporate a variety of supports, such as physical assistance, physical prompts, verbal prompts, and technology. The document *The Extensions and Adaptations of the South Carolina Curriculum Standards for Students Participating in Alternate Assessment* became the focus of the portfolio assessment process, HSAP-Alt performance tasks, and the professional development training. In 2002, this document was revised and renamed the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment*, but it was still aligned to curriculum standards for prekindergarten through grade 2. This work was based on the IDEA requirements and the thinking at the time about how students with significant cognitive disabilities should be included in the general education curriculum and assessment.

Beginning with the 2000–01 school year, students in grades 3–8 who met the participation criteria for alternate assessment were assessed with the portfolio assessment, PACT-Alt. In 2003, a high school assessment, HSAP, which was designed to meet AYP requirements, was added to the state assessment system, and an alternate to HSAP was developed to measure student proficiency in ELA and mathematics. A Stakeholder Committee with expertise in high school instruction of students with significant cognitive disabilities and academic standards was convened to guide the development of the high school alternate assessment, HSAP-Alt. The committee recommended designing an assessment based on performance on a series of tasks linked to the state curriculum standards. The HSAP-Alt consisted of a series of scripted performance tasks in ELA and mathematics with scaffolded administration and scoring procedures aligned with the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment*.

One critical piece of the development and implementation process of PACT-Alt and HSAP-Alt was the provision of intensive professional development related to standards-based instruction, much of it based on the work of Harold Kleinert and Jacqui Farmer Kearns. A resource for professional development was their book *Alternate Assessment: Measuring Outcomes and Supports for Students with Disabilities.* Professional development was essential to the implementation of the portfolio assessment because the teacher was responsible for teaching a student the content related to the academic standards, assessing the student's progress, and providing evidence of the instruction and progress in the portfolio. Prior to the implementation of the alternate assessment and the IDEA requirement to include students with disabilities in the general education curriculum, many students with disabilities, especially those with significant disabilities, and their teachers had been excluded from standards-based instruction and professional development related to academic standards.

Transition from PACT-Alt and HSAP-Alt to SC-Alt

After seeking input on the vision of a new alternate assessment on alternate achievement standards from the Advisory Committee and teachers who were conducting alternate assessment, SCDE wrote an RFP for the redesign or design of the alternate assessment system. The design was to be consistent with South Carolina's commitment to the instruction and assessment of students with significant cognitive disabilities and NCLB requirements. The focus was to be on grade-level academic standards. The new system was to address concerns related to teacher burden and time involved in assessment while supporting improved instruction based on state academic achievement standards. Extensive training for test administrators was to be integrated into the design of the assessment.

In September 2004, a contract was awarded to the American Institutes for Research (AIR) to assist the state in revising the alternate assessment. AIR managed the administration and analyses of the PACT-Alt and HSAP-Alt assessments during the 2004–05 and 2005–06 school years while developing the new alternate assessment, the South Carolina Alternate Assessment (SC-Alt), with SCDE.

American Institutes for Research

The American Institutes for Research (AIR) has more than 50 years of experience as a nonprofit organization dedicated to assessment, behavioral science, and educational research. Subcontractors for the project include Measurement Incorporated, a leader in the field of hand-scoring customized assessments and in printing, packaging, distribution, and retrieval services, and INSITE, a company with a long history of working with SCDE. AIR developed the South Carolina HSAP and the EOCEP programs and has enjoyed a successful collaboration with SCDE for a number of years.

Chapter 2: Test Development

The South Carolina academic content standards are the basis for alignment across the state for district and school curricula, classroom instruction, units of study, and learning experiences. The curriculum standards are the basis for the Palmetto Achievement Challenge Tests (PACT), the High School Assessment Program (HSAP), and the alternate assessment. An initial step in the design of the new assessment was developing Assessment Standards and Measurement Guidelines (ASMGs).

Development of the Assessment Standards and Measurement Guidelines

In April 2005, a committee comprising South Carolina special education teachers, content specialists, SCDE staff, and AIR staff designed the ASMG document to support the new assessment development. The process involved extending the state academic standards in ELA, mathematics, science, and social studies in grade bands 3–5, 6–8, and 10 to be accessible to students with significant cognitive disabilities. This document replaced the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment.*

The ASMGs are the foundation for the development of the assessment tasks for the SC-Alt. The ASMGs in each content area are distillations of the essence of South Carolina curriculum standards in each grade level.

Each content area committee reviewed the large array of standards and prioritized those in grade bands 3–5, 6–8, and 10 that they deemed most important to students "now" and "in the future." They then reduced the complexity of these standards, while retaining the essence of the grade-level content knowledge and skills, to make the academic standards appropriate and accessible for students with significant cognitive disabilities. The committee was careful to address both the depth and the breadth of the academic standards and used professional judgment based on experience with the population and the content to determine the standards to be assessed. The resulting document provides the link to the grade-level standards and indicators in the state academic standards.

The measurement guidelines give task writers and teachers the specificity necessary to translate the assessment standards into assessment tasks and items and classroom instruction. A list of individuals who were involved in this process is included in each ASMG content document.

NOTE: The ELA committee recommended that the standards in the Research Goal not be included in the assessment standards. The rationale for this recommendation was that this goal is not tested to any great extent in PACT because this content is primarily taught and assessed at the classroom level. Committee members, however, indicated that the Communication Goal included standards that they deemed very important to this population and they recommended including assessment standards for this strand.

The State Board of Education adopted revised ELA and mathematics academic standards in August 2007. Work is currently under way to align the ASMGs to the revised ELA and mathematics standards. State legislation is pending that calls for replacing the high school

physical science end-of-course assessment for all students with a biology end-of-course assessment; therefore, work is also under way to develop biology ASMGs. The adoption of these new standards occurred outside the cyclical review timetable and has a direct impact on the ongoing schedule for developing additional tasks for the task pool.

Stakeholder Input into the Development of the SC-Alt

To ensure the validity of the overall assessment process, a great deal of time and effort was spent obtaining input from various sources, including the State Alternate Assessment Advisory Committee, classroom teachers, parents, and other agency personnel.

South Carolina State Alternate Assessment Advisory Committee

The State Alternate Assessment Advisory Committee meets quarterly to provide oversight to the SC-Alt. The committee includes members of the original Alternate Assessment Steering Committee and the High School Stakeholder Committee. The committee also includes parents, special educators, representatives of higher education, content specialists, special education directors, and district test coordinators. Additional members include representatives from the Department of Disabilities and Special Needs, the University of South Carolina School of Medicine, the South Carolina Assistive Technology Project, the South Carolina Interagency Deaf-Blind Project, the Autism Society of South Carolina, and Pro-Parents of South Carolina.

The Advisory Committee provided input on its expectations for the revised alternate assessment during the first meeting with the contractor, AIR, on November 5, 2004. SCDE and AIR staff reported each step of the development process to the Advisory Committee at each meeting and sought its advice and recommendations.

Early Development Activities

At the recommendation of the Advisory Committee, AIR item writers visited classrooms in South Carolina during January and February 2005 to observe teaching strategies and materials that were in use. They also reviewed PACT-Alt portfolios for examples of evidence that teachers used to demonstrate progress toward proficiency on grade-level standards and examined the characteristics of the HSAP-Alt performance event in order to build on the existing system.

Teacher focus groups convened during January 2005 obtained feedback from teachers on the types of tasks they believed were appropriate, the protocol format they preferred, and the materials they recommended for inclusion in the assessment.

Qualified item writers employed by AIR were trained to write tasks and items specifically aligned with the alternate assessment standards and measurement guidelines. Item writing teams included AIR staff with expertise in the content areas; alternate assessment specialists; and consultants in the areas of instruction of students who are blind and visually impaired, students who are deaf and hard of hearing, and students with cognitive disabilities.

On February 14, 2006, prior to the development of science and social studies tasks, SCDE staff and the AIR alternate assessment specialist provided additional training to the writing teams. The training was based on *Designing from the Ground Floor*, materials developed by the National Alternate Assessment Center (2005).

Consideration of universal design was a focus throughout the development process. Items including passages and response options were developed to use objects, pictures, picture symbols, words, and numbers. Several tasks in all four content areas and at different levels of complexity were piloted with South Carolina teachers and students in March and May 2005. AIR staff then interviewed the pilot teachers to determine the item characteristics and parameters that teachers believed worked well or did not work.

Summary of the Development and Review of the SC-Alt Tasks

- The task and item development process began with the creation of task kernels. AIR was primarily responsible for the majority of task kernels, with input from SCDE and teachers in South Carolina. Task kernels are basic ideas for an assessment activity, stimulus materials, and purpose, which, based on their relation to the South Carolina ASMGs, were used to develop a task and its items.
- SCDE reviewed the task kernels and provided feedback to AIR on which kernels were acceptable, which were unacceptable, and which needed revision. These reviews included alignment with the ASMGs.
- AIR item writers developed the items and stimulus materials. These items were reviewed internally by the content experts for clarity, quality, and alignment with the ASMGs.
- Following the comprehensive AIR internal review, the tasks and items underwent technical review by AIR to ensure that the items were properly keyed and scaffolded, the instructions were appropriate, the stimulus materials were interpretable, and the items were generally consistent in design with other tasks and items under development.
- Items that passed internal review by the AIR development staff were reviewed by the senior content lead for each content area and the senior alternate assessment specialist. This review ensured that within the content area, tasks and items followed the design of the assessment and were consistent with respect to format, presentation, and general administration procedures.
- Before items were passed to SCDE, the project director reviewed all items to ensure that they were consistent with the foregoing factors across content areas and grade bands.
- Following the final internal AIR review, items were passed to SCDE for its review. During this process, SCDE staff, including content specialists, special educators, and assessment specialists, provided feedback to AIR on the design of the tasks and items, the alignment of items to the ASMGs, and the appropriateness of the items for use in South Carolina. Some items were revised by SCDE to improve alignment with the ASMGs.

- Approved items were placed into tasks for a small-scale tryout, conducted by AIR with the assistance of teachers in South Carolina and Northern Virginia and AIR staff. These tryouts provided invaluable information regarding the clarity of instructions, the utility of the stimulus materials, and the success of the items and tasks in producing expected responses. Items that showed obvious problems were revised or discarded.
- After changes were made to the prototypes as a result of the pilots and tryouts, a committee of South Carolina teachers was convened on July 12, 2005, to review the revised tasks and provide further input and recommendations.

Content, Bias, and Sensitivity Reviews

Once small-scale tryouts were concluded, AIR, SCDE, and educators in South Carolina reviewed the tasks and items for alignment with the ASMGs and for bias and sensitivity concerns. Committees comprising teachers of students with significant cognitive disabilities, representatives of higher education, special education administrators, experts in the instruction of students with limited English proficiency, and content experts from across the state participated in these reviews to consider the following:

- alignment to the ASMGs,
- bias for specific groups and types of disabilities,
- accessibility of the tasks to the entire population for whom the test was designed,
- specific characteristics of items that tend to exhibit bias or are not appropriate for or sensitive to the characteristics of student subgroups (e.g., exclusionary language, stereotypes),
- format and content of the tasks,
- accessibility of materials, and
- clarity of instructions and ease of administration.

The review committee meetings were conducted in November 2005 and May 2006. During the reviews, some items were recommended for revision or elimination by the committee members.

Development of Field-Test Forms

- On the basis of the feedback from all the steps above, AIR conducted a final review and sign-off for all items and tasks. Following this review, the items and tasks were affirmed ready for field-testing.
- Prior to assembling tasks into field-test forms, the senior content lead for each content area and the project director reviewed the items and tasks one last time to determine whether the revisions were appropriate and maintained the alignment of the item to the targeted assessment standard and measurement guideline.
- Tasks and their items were then placed into field-test forms consistent with the specifications described earlier.

Item Data Review

- After field-testing, AIR and SCDE staff, including alternate assessment specialists, psychometricians, content specialists, and special educators, met to review the field-test statistics.
- They reviewed the statistics associated with each item and task to determine whether the items were functioning within expectations and whether the tasks were appropriately placed within the instrument. The statistical criteria applied to the field-test item data and to the operational item data are described in Chapter 5.
- The committee also considered teacher comments on specific items from the field test, data from field-test observations, and the results of the alignment studies to make decisions about the inclusion of items in the operational assessment.
- Items that did not meet these criteria were retained for possible future operational use (or were revised for recalibration).
- The Item Data Review meetings were conducted in August 2006 and June 2007.

Development of Operational Task/Item Pool

- AIR once again reviewed all data associated with the tasks and items to determine whether the items were functioning as expected and were useful for measuring the achievement of students in South Carolina
- Items that survived all review and analysis criteria were placed into the operational task/item pool.

Design and Development of the SC-Alt Field Tests

Following the task development process, the field-test forms were designed and produced. The primary purposes of the field-test administrations for English language arts and mathematics (spring 2006), science (fall 2006), and social studies (spring 2007) were to produce data to evaluate SC-Alt tasks and items and to guide the assembly of operational test forms to be used in 2007 and beyond. Student scores based on field-test data were not reported.

The design, data collection, and analysis of the 2006 field test in ELA, mathematics, and science and of the 2007 field test in social studies were discussed in the *Spring 2007 Operational Technical Report*. This section describes the data collection, linking, and scaling in 2008 social studies field-test administrations; forms assembly for the operational tests; and the design for operational forms since 2007.

Spring 2008 Embedded Social Studies Field Test

The social studies task "George Washington" exists in two forms: a text version and a symbolate version. It is a high-complexity task for grade bands 3–5 and 6–8. The text version of the George Washington task was examined and calibrated as part of the spring 2007 social studies field test. However, the symbolate version of the George Washington task was field-tested, embedded in the spring 2008 operational administration of the social studies assessment.

It will be interesting to evaluate whether the items in both versions of the task function psychometrically and cognitively in the same manner and whether the items in both versions of

the task are equivalent in difficulty. If the items function similarly and are equivalently difficult, then the two versions of the task can be used interchangeably, and inferences from task performance are comparable. If items from the two versions function differently or differ in difficulty, then the tasks must be treated as unique, non-interchangeable tasks (e.g., must be calibrated separately).

The embedded field test presented the symbolate George Washington task as a supplementary task, added onto the 12-task operational test booklet (or administered separately). Teachers were aware that they were administering a unique task but were encouraged in training not to treat it any differently from the other tasks. Teachers were instructed to present the George Washington field test to all students, no matter whether the students were tested in the low-, medium-, or high-complexity ranges (see Exhibit 35).

Use of the Student Placement Questionnaires

The Student Placement Questionnaires (SPQs) are brief structured rating instruments that represent the range of communication levels and cognitive-academic functioning found in the population of alternate assessment examinees. AIR developed the SPQ for the South Carolina Alternate Assessment program.

The student placement process is intended to achieve several important goals:

- It matches student achievement levels with the difficulty of the tasks and items that are administered.
- It allows a maximum number of student item responses at an appropriate level of difficulty.
- It minimizes fatigue by targeting the assessment to the student.
- It supports the psychometric rigor of student scores. A student is administered a better targeted test than one that contains many items the student might find too difficult. Better test targeting contributes to better score reliability. Inasmuch as fatigue effects from the student's limited attention span are reduced, validity of the overall assessment is enhanced.

Teachers completed the SPQs in each content area to identify the most appropriate starting task for each student. For each subject, the SPQs prompted the teacher with 12 or 15 "can do" questions (e.g., can this student recognize the sun, moon, Earth?). The questions were grouped by major content standards and sampled across low-, moderate-, and high-complexity levels. Each question rated the student's functioning on a 4-point scale, valued 0 to 3. Answering the 12 or 15 questions of each SPQ, summing the total score, and identifying the most appropriate starting task in a look-up table took test administrators approximately 6 or 7 minutes.

The look-up table identified ranges of SPQ scores that corresponded to one of three starting tasks. Teachers used the SPQs to assign students to starting points on the assessment. Cut points for the science SPQ were based on the rules derived for the mathematics SPQ but were altered for the number of items on the science SPQ. Details regarding the student participation, analysis, and conclusions drawn from this field test appear below.

Administration: Placement and Stopping Rules

After teachers identified the most appropriate starting task for a student, they followed several rules as they administered the starting task and subsequent tasks. In the fall 2006 science field test, they administered (a) all items in the first four tasks and (b) as many items as possible of the three subsequent tasks—at a minimum, the first two items of each of these tasks. (For detailed placement and stopping rules for the spring 2008 operational and field-test administrations, see Appendix B.)

SPQ Summary

The preceding discussion reviewed some of the implementation procedures for the SPQ. Here we review two of the technical characteristics of the SPQ: the method used to select the SPQ recommended starting task and the usefulness of the SPQ as an indicator of student starting task.

The technical development of the SPQ and determination of the cut points to determine starting tasks are fully described in American Institutes for Research, 2008, *South Carolina Alternate Assessment (SC-Alt): Technical Report for English Language Arts and Mathematics Field Test Administration, Spring 2006*.

Usefulness of the SPQ for Determining the Starting Task. AIR gathered information from the 2008 administration regarding the agreement between SPQ recommended start points and the final observed start points as determined by a review of the 2008 item data. The results of this study are reported in detail in Chapter 8.

Use of the SPQ pre-assessment score is only the first step in the procedure used by the test administrator in determining where the student should start the assessment. The instructions for using the SPQ include procedures requiring teachers to adjust the starting point below the SPQ recommended start point when the student is not successful on the first administered task. Alternately, after reviewing the assessment, some teachers may have judged that a student needed to start at a higher level than recommended by the SPQ. This result occurred almost exclusively when the SPQ recommended Task 1 as the starting point.

The results of the 2008 study indicate that the agreement between the SPQ recommended start point and the observed start point by content area were 92% for ELA and mathematics, and 88% for science. Since the test administrator is required to make adjustments based on the student's success on the first task, and these adjustments are reflected in the agreement rates, the SPQ appears to be working very effectively for targeting the first task to begin the assessment process.

Teacher Scoring Accuracy

The design of the SC-Alt includes test administrator (teacher) scoring of student responses. The degree of accuracy with which the test administrator evaluates student performance determines whether the student receives the correct scores and the correct performance level.

A special video study was conducted during the 2008 administration to confirm that test administrators were following all scoring procedures accurately. For this study, scoring accuracy refers to the degree to which teachers follow scaffolding and scoring directions correctly and assign correct scores to student responses.

Scoring accuracy by the test administrators was evaluated by having trained raters at AIR review the videotapes of the test administrations and assume the role of the test administrator in scoring student responses. The AIR raters did not know the scores assigned by the test administrators at the time of their own assignment of scores. After the raters concluded their scoring of the student responses, the consistency between the test administrators and AIR raters was determined.

The scoring consistency analyses are presented in Appendix C, Exhibit 39. The results indicate that there was a high degree of consistency between the scoring of the test administrators and the AIR raters, suggesting that test administrators in South Carolina understood the scoring procedures and implemented them accurately when scoring student responses.

2008 Operational Test Booklets and Administration and Scoring Procedures

For each grade-band test form in each content area, tasks and items were selected that met the statistical criteria and that covered the breadth of the targeted ASMGs. The 2008 operational test forms were unchanged from the 2007 operational administration, with tasks ordered by increasing difficulty as indicated by the empirical difficulty of the first item in each task, which was determined by Item Response Theory (IRT) analysis. The goal was to use technically sound assessment instruments to support valid inferences about what students know and can do relative to the ASMGs in each content area.

The SC-Alt operational administration in spring 2008 included three sets of test materials in English language arts, mathematics, and science: one for the 3–5 grade-band assessment, one for the 6–8 grade-band assessment, and one for the grade 10 assessment. (Social studies used sets of materials for the grade band 3–5 and 6–8 assessments. Grade 10 is not part of the social studies assessment.) Teachers (test administrators) received a *Test Administration Manual* (*TAM*) and comprehensive training based on the manual and the test materials.

The 2008 test booklets contained 12 operational tasks in each content area. Social studies contained one additional 13th task for a field-test administration. Operational tasks were arranged in test forms in the order of the empirical difficulty of the first item in each task. Each test form (elementary, middle, and high school) included linking tasks to support psychometric linking of the grade-band score scales. Each task consisted of four to eight separate items. Teachers were instructed to administer a minimum of five or seven tasks to each student, depending on the SPQ designated starting point, and to continue administration of subsequent tasks until the student was no longer successful.

Teachers also received other materials with each test booklet:

- a manipulatives kit (with printed and physical manipulatives for all tasks),

- an Answer Folder for each participating student, and
- a Student Placement Questionnaire and directions for determining the starting task for each student.

Exhibit 1 summarizes the operational grade-band assessments and the numbers of tasks in each grade assessment for 2008.

Exhibit 1: Numbers of Tasks in Each Grade-Band Assessment, 2008 Operational Test

Grade Band	Total in Each Grade Band	
	ELA, Math, Science	Social Studies
3–5	12	12 (+ 1 FT task)
6–8	12	12 (+ 1 FT task)
10	12	---

The approximate test length for each grade band assessment for the 2008 administration was 60 items (12 tasks × an average 5 items per task) and 120 score points (60 items × an average 2 points per item).

Linking Tasks in Each Grade-Band Assessment

All tasks in each SC-Alt grade-band assessment align with the ASMGs in that grade band. For example, the first two items in Task 9, which is part of the grade band 3–5 ELA assessment (see Exhibit 2), align with Measurement Guideline 27, "Identify the problem and/or the solution in a story or drama," and Measurement Guideline 16, "Identify the impact of a given cause or effect on a given character." These Measurement Guidelines are linked to State Academic Standards and Indicators 3-R2.2 and 5-R1.11, respectively, from grade band 3–5.

Because adjacent grade-band score scales are linked psychometrically, some tasks in each grade-band assessment align with ASMGs in both adjacent grade bands. For example, ELA Task 10, which provides data for psychometric linking of the grade bands 3–5 and 6–8 score scales, aligns with ASMGs at both grade bands. Similarly, item 1 in Task 10 aligns with ASMG 30, "Identify the purpose of a text"; that ASMG is linked to State Academic Indicator 3-R2.10 at grade band 3–5 and State Academic Indicator 6-R2.9 at grade band 6–8 (see Exhibit 2).

All items in linking tasks are designed to be appropriate for students in *both* adjacent grade bands. The alignment studies (discussed in Chapter 8) confirm that all tasks in each grade band, including linking tasks, align with ASMGs for each separate grade band and with the corresponding grade-band academic content standards. In addition, the corresponding grade-level State Academic Standards and Indicators to which the ASMGs are linked *do* differ across the adjacent grade bands. (See Appendix A.)

Exhibit 2: Two Tasks from the Grade Band 3–5 ELA Assessment That Illustrate the Alignment of Items of Grade-Band ASMGs and State Academic Standards

Item	SC-Alt ASMG		Corresponding State Academic Standards and Indicators from Grades 3–5		Corresponding State Academic Standards and Indicators from Grades 6–8
Task 9					
Item 1	Recognize conflict in stories: ▪ 27. Identify the problem and/or the solution in a story or drama.		3-R2.2: Demonstrate the ability to identify problem and solution in a work of fiction or drama.		--
Item 2	Determine cause and effect in texts read aloud or independently: ▪ 16. Identify the impact of a given cause or effect on a given character.	☐	5-R1.11: Demonstrate the ability to analyze cause and effect.		
Item 3	Analyze plots, characters, and settings in literature: ▪ 25. Identify and describe characters, settings, and events in a story.		3-R2.1: Demonstrate the ability to analyze characters, setting, and plot in a literary work.		
Item 4					
Item 5	Make predictions about text: ▪ 9. Use pictures and words to make predictions about texts read aloud or independently.		3-R1.7): Demonstrate the ability to make predictions about stories.		
Task 10					
Item 1	Identify the author's purpose: ▪ 30. Identify the purpose of a text.		3-R2.10: Continue identifying the author's purpose in a variety of texts.		6-R2.9: Demonstrate the ability to identify the author's purpose in texts in a variety of genres.
Item 2	Recall details: ▪ 7. Recall details in		3-R1.5: Demonstrate the ability to recall		6-R1.5: Demonstrate the ability to identify

Item 3	tests read aloud or independently. Follow written directions: ▪ 17. Follow written one-step or multistep directions (presented in words/pictures/symbols/ objects).		details in texts. 3-R1.13: Demonstrate the ability to follow a logical sequence of written directions to complete a task.		the details that support the thesis of a particular text. 6-R1.11: Demonstrate the ability to follow multistep directions such as those for preparing applications and completing forms.
Item 4					
Item 5					
Item 6	Make predictions about text: ▪ 9. Use pictures and words to make predictions about texts read aloud or independently.		3-R1.7: Demonstrate the ability to make predictions about stories.		6-R1.6: Demonstrate the ability to make predictions about stories.

Chapter 3: Spring 2008 Operational Test Administration

This section describes the spring 2008 operational test administration:

- student participation for the spring 2008 administration;
- demographics of participating students;
- test administration window, materials, and timelines;
- test administrator requirements;
- test administrator training;
- pre-assessment using the Student Placement Questionnaire;
- fidelity of administration and accuracy of scoring; and
- test security provisions.

Student Participation for the Spring 2008 Administration

Students participating in the spring 2008 operational administration were those students whose IEP team had determined that they met the following SC-Alt participation criteria for alternate assessment and who were ages 8–13 or 15 on September 1, 2007. These are the ages of typical students who are in grades 3–8 and 10.

- The student demonstrates a significant cognitive disability and adaptive skills, which result in performance that is substantially below grade-level achievement; expectations even with the use of accommodations and modifications.
- The student accesses the state-approved curriculum standards at less-complex levels and with extensively modified instruction.
- The student has current adaptive skills requiring extensive direct instruction and practice in multiple settings to accomplish the application and transfer of skills necessary for application in school, work, home, and community environments.
- The student is unable to apply or use academic skills across natural settings when instructed solely or primarily through classroom instruction.
- The student's inability to achieve the state grade-level achievement expectations is not the result of excessive or extended absences or social, cultural, or economic differences.

Exhibit 3 indicates the age ranges of students who participated in the SC-Alt in spring 2008. Exhibit 4 indicates the alternate assessment eligibility categories that were placed in each eligible student's state precoding file. (Precoding files enabled SCDE and AIR to ensure that the appropriate SC-Alt assessment materials were delivered to teachers in time for the spring 2008 administration.)

Exhibit 3: Age Reference Sheet for 2007–08 Alternate Assessment, Spring 2008 Operational Administration

	Corresponding Birth Date Range			
Age as of 9/1/07	**Beginning DOB**	**Ending DOB**	**Test Required 2007–08**	**Precode AA Eligibility Code**
3	09/02/03	09/01/04	none	
4	09/02/02	09/01/03	none	
5	09/02/01	09/01/02	SCRA-Alt[1]	1
6	09/02/00	09/01/01	SCRA-Alt	1
7	09/02/99	09/01/00	none	5
8	09/02/98	09/01/99	SC-Alt Elementary	2
9	09/02/97	09/01/98	SC-Alt Elementary	2
10	09/02/96	09/01/97	SC-Alt Elementary	2
11	09/02/95	09/01/96	SC-Alt Middle	3
12	09/02/94	09/01/95	SC-Alt Middle	3
13	09/02/93	09/01/94	SC-Alt Middle	3
14	09/02/92	09/01/93	none	5
15	09/02/91	09/01/92	SC-Alt HS	4
16	09/02/90	09/01/91	none	5
17	09/02/89	09/01/90	none	5
18	09/02/88	09/01/89	none	5
19	09/02/87	09/01/88	none	5
20	09/02/86	09/01/87	none	5
21	09/02/85	09/01/86	none	5

Exhibit 4: Precode Project Coding (Alternate Assessment Eligibility Field)

Code	SASI Dropdown List Description	Full Description
0	Criteria not met	The student does not meet criteria for alternate assessment.
1	**SCRA-Alt**	The student requires alternate assessment and meets the age eligibility requirement for **SCRA-Alt** this current school year (5 years old for kindergarten or 6 years old for first grade on September 1, 2007).
2	**SC-Alt Elem School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt Elem**

[1] SCRA-Alt: South Carolina Readiness Assessment—Alternate: an assessment of student readiness administered to students in kindergarten and first grade. The SCRA-Alt is a teacher rating scale and is not included in the state accountability system.

		School form this current school year (8-10 years old on September 1, 2007).
3	**SC-Alt Middle School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt Middle School form** this current school year (11-13 years old on September 1, 2007).
4	**SC-Alt High School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt High School form** this current school year (15 years old on September 1, 2007).
5	**AltAssess NotAgeElig**	The student requires alternate assessment, **but does not meet the age eligibility requirements to be assessed with SCRA-Alt, or SC-Alt this current school year** (i.e., the student was age 7, 14, 16, or older than 16 on September 1, 2007).

Demographics of Participating Students

This section describes the demographics of participating students by test form (elementary, middle, or high school). Exhibit 5 presents the student demographics for participating students in each grade band.

For the purpose of this report, the inclusion of students was based on the same criteria applied in the reporting of student scores. A student was included if the following criteria were met: (1) a signed security affidavit was received for the student, (2) the student was not noted to be excluded from reporting for some other reason (e.g., inappropriate administration procedures), and (3) the number of coded responses met the attemptedness requirement for student scoring (i.e., five valid responses) in at least one content area. The population of students reported, therefore, includes 1,232 elementary school test forms, 1,052 middle school test forms, and 373 high school test forms.

According to the attemptedness requirements, a student's responses to a test form could be assigned to one of three completion status categories: completion ("student satisfied attemptedness rule"), invalid ("student did not satisfy attemptedness rule"), or not tested ("student did not answer any content area items"). For all content areas, the majority of students reported completed the administered test form; 98% or more of all students completed ELA and math while 69% to 72% completed science and social studies.[2] Of the remaining student records, typically 1% or less of reported test forms were categorized as not tested or not meeting the attemptedness criteria.

Given that the number of students assessed by the high school test form was approximately one-third the number of students assessed by either the elementary or the middle school forms, the proportion of demographic characteristics of the student population was relatively consistent across grade bands. In terms of ethnicity, African American students made up at least 50% of the assessed students across grade bands (51% to 55%); Hispanic students accounted for approximately 3% (1% to 4%); and White students accounted for 40% to 42% of

[2] Not all students were required to complete the science and social studies subject areas.

the students across grade bands. Other ethnicities each accounted for less than 4% of the assessed population. Gender was also relatively consistent across grade bands, showing a slight proportional decrease of males from elementary to middle and then high school, with approximately a two-to-one ratio of male students (67% to 63%) to females (33% to 37%).

The classification of students in terms of English language proficiency was also consistent across grade bands. The majority of students (97% to 98%) were classified as "English Speaker II," meaning that they had never been coded as an ESL student. The remaining language proficiency classifications each accounted for less than 1% of students by grade band with the exception of "Pre-functional" (1% to 3%), indicating that the student scored "pre-functional" on the English language proficiency assessment and is receiving English as a second language (ESL) services. The percentage of pre-functional ESL students decreased across grade bands.

The grade reported for a student in the school's database is the grade reported for funding purposes (EFA grade) and is often determined by the location of the student's educational program instead of by the student's age or years in school. Therefore, approximately 9% of students administered the elementary form (for students ages 8–10, the typical ages of students in grades 3–5) had reported EFA grades lower than grade 3 or higher than grade 5, with most of these students classified in the contiguous grades of 2 and 6. Of students administered the middle school form (for students ages 11–13, the typical ages for grades 6–8), 24% of the students were reported at grades below grade 6 or above grade 8. The vast majority of these students were classified as grade 5 students (17% of all middle school form students), which indicates that these students were being served in educational programs housed in elementary schools. Of the students administered the high school form (for students age 15), 75% were reported as grade 9 or grade 10 (36% and 39%, respectively). Sixteen percent (16%) of the high school form students were reported as grade 8 students, indicating that these students were being served in educational programs housed in middle schools. The purpose of assigning SC-Alt grade band forms by age is to ensure that students are instructed and assessed on the appropriate grade band curricula regardless of where their educational programs are housed.

Fourteen different disability codes were reported for students assessed with the SC-Alt. The coding system allowed students to be coded with more than one disability code. Students with the primary disabilities of moderate mental disability, mild mental disability, severe mental disability, and autism made up 88% to 91% of the students assessed with the SC-Alt (88%, 91%, and 88% for the elementary, middle school, and high school forms respectively). Of these four primary disabilities classifications, the largest SC-Alt participation group was moderate mental disability (30-49%), followed by mild mental disability (18-27%), autism (12-21%), and severe mental disability (9-11%). Although a few students would have a primary disability of speech or language impairment, the vast majority of students received this code because they were receiving speech/language therapy as a supplementary service.

Exhibit 5: Summary of Demographic Information

	Elementary		Middle		High	
	N	%	N	%	N	%
STUDENT'S ETHNICITY						
Blank	1	0.08	.	0	2	0.54
African American	647	52.52	538	51.14	205	54.96
African American/American Indian	3	0.24	.	0	1	0.27
American Indian	5	0.41	3	0.29	.	0
Asian	8	0.65	12	1.14	2	0.54
Hawaiian/Pacific Islander	1	0.08	1	0.10	1	0.27
Hispanic	47	3.81	35	3.33	3	0.80
Other	3	0.24	4	0.38	.	0
White	494	40.10	446	42.40	158	42.36
White/African American	18	1.46	11	1.05	1	0.27
White/American Indian	.	0	2	0.19	.	0
White/Asian	5	0.41	.	0	.	0
STUDENT'S GENDER						
Blank	1	0.08	.	0	2	0.54
Female	405	32.87	356	33.84	137	36.73
Male	826	67.05	696	66.16	234	62.73
ESL (LANGUAGE)						
Blank	6	0.49	2	0.19	7	1.88
Beginner	1	0.08	1	0.10	.	0
Beginner Waiver	.	0	1	0.10	.	0
English Speaker I	.	0	1	0.10	.	0
English Speaker II	1192	96.75	1021	97.05	364	97.59
Initially English Proficient	1	0.08	1	0.10	.	0
Intermediate	.	0	2	0.19	.	0
Pre-functional	31	2.52	23	2.19	2	0.54
Pre-functional Waiver	1	0.08	.	0	.	0
ELIGIBLE FOR FREE OR REDUCED-PRICE LUNCH						
Free	745	60.47	629	59.79	225	60.32
No	387	31.41	335	31.84	113	30.29
Reduced	97	7.87	88	8.37	30	8.04
EFA GRADE (REPORTED GRADE FOR FUNDING)						
1	4	0.32	.	0	.	0
2	65	5.28	.	0	.	0
3	454	36.85	6	0.57	.	0
4	396	32.14	23	2.19	.	0
5	271	22.00	183	17.40	9	2.41
6	29	2.35	342	32.51	2	0.54
7	1	0.08	298	28.33	10	2.68

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
8	4	0.32	164	15.59	58	15.55
9	1	0.08	22	2.09	136	36.46
10	1	0.08	8	0.76	145	38.87
11	2	0.16	5	0.48	10	2.68
12	1	0.08	.	0	1	0.27
COMPLETION STATUS: Student satisfied attemptedness rule						
ELA	1231	99.92	1048	99.62	373	100.00
Math	1228	99.68	1043	99.14	372	99.73
Science	881	71.51	759	72.15	370	99.20
Social Studies	847	68.75	723	68.73	.	0
COMPLETION STATUS: Student did not answer any content-area items						
ELA	1	0.08	3	0.29	.	0
Math	3	0.24	8	0.76	1	0.27
Science	344	27.92	291	27.66	3	0.80
Social Studies	378	30.68	324	30.80	373	100.00
Migrant Status	.	0	1	0.10	.	0
Home Schooled	.	0	1	0.10	.	0
Medical Homebound	12	0.97	23	2.19	2	0.54
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)						
Severe Mental Disability	128	10.39	111	10.55	35	9.38
Moderate Mental Disability	371	30.11	440	41.83	181	48.53
Mild Mental Disability	327	26.54	248	23.57	66	17.69
Autism	253	20.54	154	14.64	45	12.06
Deaf/Blindness	3	0.24	3	0.29	4	1.07
Emotional Disability	7	0.57	4	0.38	.	0
Hearing Impaired	72	5.84	55	5.23	16	4.29
Learning Disability	10	0.81	6	0.57	3	0.80
Multiple-Disability	83	6.74	84	7.98	25	6.70
Other Health Impaired	59	4.79	43	4.09	15	4.02
Orthopedically Impaired	59	4.79	51	4.85	11	2.95
Speech or Language Impaired	888	72.08	559	53.14	135	36.19
Traumatic Brain Injury	3	0.24	5	0.48	4	1.07
Visually Impaired	49	3.98	32	3.04	7	1.88
TOTAL	**1232**	**100**	**1052**	**100**	**373**	**100**

Test Administration Window, Materials, and Timelines

The following list presents important dates for the spring 2008 administration of the SC-Alt:

- SC-Alt test administration training for teachers new to the SC-Alt operational administration (did not administer in 2007) – four regional SCDE workshops: January 15-18, 2008
- District-level SC-Alt test administration training for all test administrators: January 28 – February 22, 2008
- Test materials arrived in district: week of February 22, 2008
- Assessment window: March 3–April 18, 2008
- Teachers returned materials to DTC-Alt: April 21–22, 2008
- Materials shipped to contractor: April 25, 2008

Teachers had approximately seven weeks to review the materials and complete the test administration. Teachers received both printed and physical manipulatives to use during test administration. They were also responsible for collecting a few common classroom items that were familiar to the student to use with several tasks.

Test Administrator Requirements

Test administrators are required to receive training on all phases of the administration of the SC-Alt and must be one of the following:

- a certified employee of the district;
- an employee of the district who is a critical needs teacher and has a letter of eligibility, an interim certificate, or a critical needs certificate;
- a substitute teacher who is certified and employed by the district on an as-needed basis;
- someone who was a certified teacher but has allowed the teaching certificate to expire owing to retirement, change of career, or some other reason and has been approved by the district test coordinator or the DTC-Alt as a qualified test administrator; or
- someone who is not certified but has been employed by the school district in an instructional capacity and has been approved by the DTC-Alt as a qualified test administrator.

If a test is administered in a location other than the school, the test administrator must meet the criteria specified above.

Test Administrator Training

Test administration training is required for all test administrators. The SC-Alt is individually administered with a standard script and scored by the test administrator as the assessment is conducted. Fidelity of administration and scoring is essential to the validity of the assessment results.

Teachers who administered the SC-Alt during the spring of 2008 who did not administer the SC-Alt in spring 2007 were required to attend a South Carolina Department of Education (SCDE) training session. In addition, all teachers who administered the SC-Alt in spring 2008, including those who attended the SCDE workshops, were required to attend a district-level SC-Alt administration training session conducted by the district test coordinator for alternate assessment (DTC-Alt). At the completion of the training sessions, each test administrator was required to sign and submit to SCDE an acknowledgment of receiving training and readiness to conduct the assessment.

The training included the following elements:

- review of the eligibility criteria for students participating in the alternate assessment;
- overview of the ASMGs, emphasizing the link to the general education standards;
- explanation of how the assessment was developed, including the role of the review committees;
- review of test administrator requirements, test security, and test materials;
- training and practice in pre-assessment using the SPQ;
- description of the assessment format and procedures:
    - setup,
    - script,
    - scoring,
    - adaptive instructions;
- making SC-Alt tasks accessible;
- overview of assistive technology and the alternate assessment;
- administration and scoring instruction and practice using released test items provided on video clips of South Carolina teachers actually administering a task to students representing a variety of disabilities and ethnicities;
- scoring qualifying round; and
- review of procedures for receiving and shipping materials back to the DTC-Alt.

Pre-Assessment Using the Student Placement Questionnaire

As noted earlier in this Technical Report, the administration of the SC-Alt uses the SPQ as a pre-assessment instrument to determine the most appropriate starting point in the assessment. Recall that the SPQ requires the teacher to evaluate the student on 12 to 15 "can do" statements addressing the student's skills and knowledge in each content area on the basis of the teacher's prior instructional knowledge of the student. A total score computed from the teacher's SPQ responses indicates the initial starting task for the assessment. Once the assessment has begun, the test administrator is required to adjust the starting point for the student if the student is not successful on the first task. Rules have been established for adjusting the starting tasks and for determining when the assessment should be concluded. The starting and stopping rules used with the SPQs for the 2008 administration are presented in Appendix B.

Fidelity of Administration and Accuracy of Scoring

This section describes the steps taken to ensure the fidelity of administration and the accuracy of scoring.

During the assessment administration, a monitor must be present to observe all assessment sessions and verify the use of proper assessment procedures and the authenticity of student responses. Monitors must be trained and sign a Test Administrator Security Affidavit to verify that the appropriate procedures were used. The Test Administrator Security Affidavit contains a coded label to link the document to the student answer folder and includes the principal's verification of the use of appropriate assessment and scoring procedures. If this document is missing, the administration is considered an invalid administration.

AIR and SCDE conducted an audit of the spring 2008 administration and scoring by requiring school system staff to videotape a sample of SC-Alt administrations. A sample of students were identified for videotaping according to the following:

- All districts implementing the SC-Alt were required to videotape at least one student administration (all content areas).
- Each teacher included in the sample was required to videotape only one student.
- The total number of videotaped administrations per district was based on the number of teachers involved in the assessment for each district.

The test administrators of the sampled students were notified of their inclusion in the sample and were given instructions for completing the videotaping. Approximately 10% of all assessed students and 33% of all test administrators were sampled. The videotapes were returned to the contractor and scored by trained raters. Ten percent of these videos were also scored by AIR's senior alternate assessment specialist. More detailed information on this study is presented in Appendix C.

Test Security Provisions

This section describes the test security procedures associated with the SC-Alt. SCDE has the following test security measures in place:

- Each local school board must develop and adopt a district test security policy. The policy must provide for the security of the materials during testing and the storage of all secure tests and test materials before, during, and after testing. Before and after testing, all materials must be stored at a location(s) in the district under lock and key.
- Each District Superintendent must designate annually one individual in each district for each mandated assessment who will be the sole individual in the district authorized to procure test instruments that are used in testing programs administered by or through the State Board of Education. The designated individual for alternate assessment is the district test coordinator for Alternate Assessment (DTC-Alt). The DTC-Alt is responsible for receiving and distributing all SC-Alt materials and ensuring that all SC-Alt administration procedures and requirements are met.
- All school and district personnel who may have access to SC-Alt test materials or to the location in which the materials are securely stored must sign the Agreement to

Maintain Test Security and Confidentiality before they are given access to the materials.

- Test administrators must be trained annually to administer the SC-Alt and must meet all test administrator requirements.
- An assessment monitor must observe all assessment sessions and verify the use of proper assessment procedures and the authenticity of student responses for each completed assessment.
- Test administrators must complete an SC-Alt Test Administrator Security Affidavit for each student they assess and submit the form with the student's assessment materials.

Chapter 4: Setting Performance Standards

This chapter describes the methods used to set standards on the SC-Alt assessments for the 2006–07 school year and beyond. It includes descriptions of achievement levels, the procedure used to set standards for each content area, the goals of the process, the composition of the panels, the workshop mechanics, and the standards set for each content area, including student impact information. Complete details of this procedure can be found in a separate report (*South Carolina Alternate Assessment Spring 2007 Standard Setting Technical Report*).

From June 25 to June 27, 2007, AIR convened a diverse panel of 105 educators and non-educators to recommend status performance standards based on the spring 2007 operational test administration data for ELA, mathematics, and science and field-test data for social studies.

Using the Item Descriptor (ID) Matching method (see Cizek & Bunch, 2007; Ferrara, Perie, & Johnson, in press), the panelists reviewed test items and the corresponding Descriptions of Achievement Levels (DALs) and then recommended performance standards for Level 2, Level 3, and Level 4 achievement levels. These standards were then translated into cut points on the student proficiency scale by AIR psychometricians. This section describes the process and outcomes of the standard-setting workshop.

Descriptions of Achievement Levels

DALs are key elements in standard-setting processes. DALs define the content area knowledge, skills, and processes that examinees at a performance level are expected to possess. The descriptions of Level 1, Level 2, Level 3, and Level 4 performance that SCDE developed make up the public statement about what and how much South Carolina educators want students to know and be able to do for each grade level and content area. In the ID Matching standard setting for the SC-Alt tests, panelists based their judgments on the DALs presented in Appendix D when they placed their cut scores.

The development of the DALs followed a multistep process involving AIR staff and SCDE staff working with committees of teachers, parents, and special education administrators. The process was begun by examining the DALS used with the other South Carolina assessment programs (PACT, HSAP, PACT-Alt, and HSAP-Alt) and the performance-level descriptors for alternate assessments used by other states. Following this preliminary study, a decision was made to draft four levels of descriptors, Levels 1–4, with the intent of having Level 3 represent "proficient performance" for NCLB reporting.

In the next step, staff examined item locations on the vertical scale for each grade band and looked for clustering of content strands and other response demands from the 2006 field tests in ELA, mathematics, and science. SCDE and AIR staff drafted DALS around these clusters as a starting point only. They then refined these drafts to ensure reference to all content strands and articulation within and across levels prior to submitting them for committee input.

A stakeholder committee met on March 30, 2007, and was charged with determining what proficiency "looked like" for students participating in each grade band of the SC-Alt. The committee reviewed the ASMGs, the draft labels for the DALs, and the draft DALs to develop refined DALs to recommend to SCDE. Members were instructed to consider descriptions that

allowed room for growth within grade bands and across grade bands and to recommend DALs that reflected an expectation that students will grow and achieve from year to year and demonstrate more knowledge. They were requested to design DALs to allow room for higher achievement.

After the meeting, SCDE and AIR staff reviewed these drafts to ensure consistency across grade bands and performance levels. The committee reviewed the DALs again on May 16, 2007, to refine them before the standard-setting workshop in June. Some additional refinement occurred during the standard-setting workshop. The official DALs were presented to the State Board of Education on September 12 and are posted on the SCDE website.

The ID Matching Standard-Setting Process

The ID Matching standard-setting process, described in the standard-setting plan submitted to SCDE and reviewed by the South Carolina Technical Advisory Committee, was used at a workshop in Columbia from June 25 to June 27, 2007, with a panel of 105 members. The panel was divided into four groups: an ELA group, a mathematics group, a science group, and a social studies group. Three tables in each workshop were assigned to anchor grade band 3–5. The other three tables were assigned to anchor grade band 10 (except in the social studies panel, which split the panel between grade bands 3–5 and 6–8). AIR staff provided training and led the participants through two rounds of ID Matching to first set the Level 3 standard and then the Level 2 and the Level 4 standards.

Before the participants made each of their recommendations using the ID Matching procedure, they were given a readiness form to ensure that they fully understood the task and were prepared to place the performance standard. Analysis of these evaluations showed unanimous agreement from the participants that they understood the task and were prepared to make performance standard recommendations.

Goals of the Standard Setting

The goals of the meeting, as stated to the panelists, were as follows:

- recommend performance standards on the ELA, mathematics, science, and social studies assessments that correspond to the DALs for Level 2, Level 3, and Level 4 performance levels;
- consider the agreement and impact data to guide judgments about item difficulty and placement of the performance standards; and
- recommend to SCDE the appropriate placement of cut points on the student proficiency scales for each grade-band assessment.

Panel Composition

The 105 panelists participated in recommending performance standards across four content areas: ELA, mathematics, science, and social studies. The overall composition of the panel followed the SCDE-provided specifications and was broadly designed to ensure that the panel was widely diverse and represented a cross-section of South Carolina's educators and non-educators. The composition of all panels is shown in Exhibit 6. The demographic breakdown of the standard-setting panelists appears in Exhibit 7.

Exhibit 6: Composition of the Standard-Setting Panels

	Panelist Role					
	Grades 3–5 Subpanel			Grade 10 Subpanel[1]		
Panel	**Table 1**	**Table 2**	**Table 3**	**Table 1**	**Table 2**	**Table 3**
ELA	Teacher - Special Education*	Administrator - Special Education*	Administrator - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*
	Curriculum Specialist-ELA	Curriculum Specialist-ELA	Administrator - Special Education	Curriculum Specialist – ELA	Administrator - Special Education	Curriculum Specialist – ELA
	Higher Education	Teacher - Special Education	Curriculum Specialist - ESOL	Teacher - Special Education	Curriculum Specialist – ELA	Curriculum Specialist - ELA
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education
	Teacher - Special Education		Teacher - Special Education		Teacher - Special Education	
Mathematics	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Administrator - Special Education*	Administrator - DTC*
	Teacher - Special Education	Administrator - Special Education	Curriculum Specialist - mathematics	Teacher - ESOL	Curriculum Specialist - mathematics	Curriculum Specialist - Autism
	Curriculum Specialist - mathematics	Curriculum Specialist - mathematics	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Curriculum Specialist - mathematics
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education
		Teacher - Special Education	Administrator –Principal	Curriculum Specialist - mathematics	Higher Education	Teacher - Special Education
Science	Teacher - Special Education*	Administrator - Special Education*	Administrator - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*
	Curriculum Specialist - science	Curriculum Specialist - science	Administrator - Special Education	Curriculum Specialist - science	Administrator - Special Education	Administrator - ESOL
	Teacher - Special Education	Teacher - Special Education	Curriculum Specialist - science	Teacher - Special Education	Curriculum Specialist - science	Curriculum Specialist - science
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education
	Teacher - Special Education		Teacher - Special Education			
Social Studies	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Administrator - Special Education*
	Curriculum Specialist - SS	Administrator –Principal	Curriculum Specialist – Social Studies	Administrator - Special Education	Teacher - Special Education	Teacher - Special Education
	Teacher - Special Education	Curriculum Specialist – Social Studies	Parent	Teacher - Special Education	Curriculum Specialist – Social Studies	Teacher - Special Education
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	

Notes (1) Grades 6–8 for Social Studies. (*) Denotes table leader.

Exhibit 7: Demographic Breakdown of Standard-Setting Panelists

		N
TOTAL		105
Gender	Female	94
	Male	11
School District	Richland 1	14
	Richland 2	10
	Horry County	5
	Lexington 2	5
	Dorchester County	4
	Florence 1	4
	Lexington 5	4
	Lexington 1	3
	Berkeley County	2
	Greenville County	4
	Charleston County	4
	Greenwood 50	2
	Kershaw County	2
	SC School for the Deaf and Blind	2
	Sumter 17	2
	Union County	2
	York 3	2
	Other School Districts (1 each)	32
Other	Department of Disabilities	1
	Pro Parents of SC	1
Race/Ethnicity	African American	19
	Hispanic	2
	White	80
	Unknown/Other	4
Position	Special Education Teacher	61
	Curriculum Specialist	23
	Special Education Administrator	12
	Administrator	3
	Higher Education	2
	ESL Teacher/Curriculum Specialist	3
	Parent*	4

*Three parents were also special education teachers and have been counted in both categories.

Standard-Setting Workshop Activities

Workshop participants recommended performance standards for the assessments during two rounds of deliberation for each DAL in each content area and in each grade band as follows.

- Set standards in anchor grade bands (3–5 and 10)
  - Participants complete Rounds 1 and 2 for each performance-level standard.
  - Table leaders *articulate* standards across grades and content areas (align them on the basis of content considerations).
- Set standards in intermediate grade band (6–8)
  - Participants complete Rounds 1 and 2 for each performance-level standard.
  - Table leaders *articulate* standards across grades and content areas (align them on the basis of content considerations and consistency with anchor grade standards).

The workshop agenda shown in Exhibit 8 shows the sequence of events for the three-day meeting.

Exhibit 8: Standard-Setting Workshop Agenda

Day	Approx. Times	Primary Activity	ELA	Mathematics	Science	Social Studies
1	8:00–11:00	Table leader training	24 table leaders (6 from each content area)			
	11:00–12:00 1:00–3:00	Panelist training and practice	105 panelists (23 to 29 from each content area)			
	3:00–5:00	Anchor grades, Level 3, Level 2, Level 4, round 1	14 panelists for grades 3–5; 13 panelists for grade 10	14 panelists for grades 3–5; 15 panelists for grade 10	14 panelists for grades 3–5; 12 panelists for grade 10	12 panelists for grades 3–5; 11 panelists for grades 6–8
2	8:00–9:00	Review Day 1, Finalize Round 1				
	9:00–11:00	Anchor grades, Level 3, Level 2, Level 4, round 2	14, 13	14, 15	14, 12	12, 11
	11:30–1:00	Anchor grades, moderation by table leaders, all content areas	24 table leaders			
	2:00–5:00	Grades 6–8, Level 3, Level 2, and Level 4, round 1	27	29	26	n/a

Day	Approx. Times	Primary Activity	ELA	Mathematics	Science	Social Studies
	2:00–5:00	Closing, final results, and evaluation for social studies	n/a	n/a	n/a	23
3	9:00–11:00	Grades 6–8, Level 3, Level 2, and Level 4, round 2	27	29	26	n/a
	11:30–12:00; 1:00–2:00	Final moderation by table leaders, all content areas	18 remaining table leaders			
	1:00–3:00	Closing, final results, and evaluation	82 remaining panelists			

Throughout the week, the panelists had many opportunities to reflect on the pattern of performance standards they were recommending. Their general conclusion was that they were satisfied that the standards made sense from a content and experiential point of view. They felt that the patterns reflected the requirements of the content standards and the realities of student performance.

With few exceptions, panelists recommended standards that followed an orderly progression of increasing achievement across levels and grade bands. Specifically, with the exception of mathematics at the grade band 6–8 and grade 10, all recommended achievement-level standards increased in difficulty in subsequent grade bands. This fact is evident by examining the scale scores associated with each recommended cut score at each grade and achievement level. Exhibit 9 through Exhibit 12 show the scale score associated with the cut score recommended by each panel. These results were achieved through the process of setting cut scores at anchor grades, making sure they resulted in consistent expectations across grade bands, and providing articulated standards as a starting point for intermediate grade bands.

Cut Score Review and the Setting of Final Cut Scores

The results of the standard-setting workshops were presented to the Technical Advisory Committee (TAC) of the Office of Assessment, SCDE, on July 27, 2007. The TAC discussed the results of the standard-setting workshops, reviewed the articulation of the cut scores by grade level, and recommended strategies to the Office of Assessment staff for improving the articulation of the final scores while respecting and maintaining the basic cut score decisions made by the workshop panelists.

The Office of Assessment staff presented the following information to the TAC for review and discussion:

- demographic and disability characteristics of the spring 2007 test participants;
- spring 2007 standard-setting results from the standard-setting workshop conducted by AIR;

- score distributions and descriptive statistics; and
- collateral data, including results from PACT, PACT-Alt, HSAP-Alt, and HSAP.

In addition, Special Education Unit staff of the Office of Assessment stated two assumptions about the students assessed with the SC-Alt and their current levels of academic instruction:

- The vast majority of students with significant cognitive disabilities will improve in their academic achievement as a result of instruction, which will result in increased achievement performance across grades (i.e., from elementary through high school).
- Many teachers of students with significant cognitive disabilities are just beginning to implement academic standards-based instruction, and therefore the students assessed with the SC-Alt in spring 2007 have not received the level of instruction that is desired or expected in the future.

A consideration of the initial cut scores in light of these assumptions identified a need to improve the articulation of cut scores across grade levels, both in terms of the scale score growth expectations on the vertical achievement scale and the percentage of students identified in each achievement level.

The TAC recommended that SCDE staff consider adjustments to improve cut score articulation to be more consistent with expectations related to the design of the test and the achievement performance of the students. The TAC recommended that careful and thoughtful adjustments to the cut scores, guided by the standard error confidence intervals around each original cut, would be reasonable and acceptable policy adjustments.

A committee of Office of Assessment staff examined the scale score articulation and the percentage of students in performance levels by grade and recommended minor adjustments to the original cuts made by the workshop panelists. The adjustments made to each cut score and the resulting final cut scores are presented in Exhibit 9 through Exhibit 12.

The standard error of the panelist-recommended cut score in Exhibit 9 through Exhibit 16 was based on estimates of the standard error of the median suggested by Huynh (2003). However, two additional details about the standard errors of the median are important to note: First, the standard errors were based on the actual recommended cut scores, and any post hoc adjustment to the cut scores was treated as a constant adjustment. In other words, the adjusted cut score still had the same standard error. Second, the standard errors were initially calculated as standard errors of the page numbers in the ordered-item booklet. In other words, a standard error of the median equal to 2 meant that the error in the panelists' recommended cut score was about plus or minus two pages. The standard error was then transformed to the scaled score metric through linear interpolation. This was possible because each page number in the ordered-item booklet has a scale score associated with it.

Exhibit 9: Panel Recommended and Adjusted Final Cut Scores—ELA

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Grades 3–5					
Level 2	403	13.75	None	403	2.96
Level 3	466	9.54	None	466	1.59
Level 4	491	12.26	None	491	1.73
Grades 6–8					
Level 2	417	9.64	None	417	3.81
Level 3	473	7.99	0.5	477	1.09
Level 4	501	9.18	None	501	1.45
Grade 10					
Level 2	429	10.56	None	429	3.38
Level 3	478	9.11	1	487	0.66
Level 4	503	9.68	1	514	1.77

Exhibit 10: Panel Recommended and Adjusted Final Cut Scores—Mathematics

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Grades 3–5					
Level 2	423	10.22	-1	413	0.66
Level 3	476	9.59	None	476	0.21
Level 4	526	14.48	None	526	4.63
Grades 6–8					
Level 2	425	10.18	None	425	0.50
Level 3	476	9	1.5	489	0.16
Level 4	529	10.46	0.5	534	0.74
Grade 10					
Level 2	434	11.93	None	434	2.19
Level 3	476	14.76	1.5	498	1.97
Level 4	528	13.19	1	541	3.82

Exhibit 11: Panel Recommended and Adjusted Final Cut Scores—Science

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Grades 3–5					
Level 2	430	10.83	None	430	1.51
Level 3	474	10.36	-0.5	469	3.25
Level 4	496	10.38	None	496	0.81
Grades 6–8					
Level 2	447	9.66	None	447	0.06
Level 3	484	9.61	0.5	489	0.50
Level 4	514	11.33	None	514	0.95

Grades 10					
Level 2	463	11.72	None	463	4.71
Level 3	492	14.44	1	506	8.13
Level 4	535	14.78	None	535	1.45

Exhibit 12: Panel Recommended and Adjusted Final Cut Scores—Social Studies

	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
Performance Level	**Scale Score**	**Scale Score Standard Error of Measurement (± SEM)**	**Level of Adjustment (± SEM)**	**Final Cut Scale Score**	**Standard Error of Cut Scale Score**
Grades 3–5					
Level 2	423	16.64	None	423	2.98
Level 3	485	14.39	0.5	492	11.93
Level 4	549	14	None	549	2.04
Grades 6–8					
Level 2	439	14.04	None	439	5.96
Level 3	490	12.58	1.5	503	1.28
Level 4	560	26.91	None	560	10.57

The final cut scores, the percentage of students performing at each performance level, and the cumulative percentage of students at or above each level are presented in Exhibit 13 through 16. The final cut scores were approved by the State Superintendent of Education and were presented to the South Carolina State Board of Education on September 12, 2007.

Exhibit 13: Percentage of Students at Each Performance Level—ELA

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	12.6 %	100.0 %
Level 2	403	25.4 %	87.4 %
Level 3	466	21.9 %	62.0 %
Level 4	491	40.1 %	40.1%

Grades 6–8			
Level 1	—	12.9 %	100.0 %
Level 2	417	23.3 %	87.2 %
Level 3	477	14.9 %	63.9 %
Level 4	501	49.0 %	49.0 %
Grade 10			
Level 1	—	13.4 %	100.0 %
Level 2	429	23.6 %	86.6 %
Level 3	487	12.5 %	63.1%
Level 4	514	50.6 %	50.6 %

Exhibit 14: Percentage of Students at Each Performance Level—Mathematics

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	14.3 %	100.0 %
Level 2	413	30.8 %	85.7 %
Level 3	476	29.3 %	54.9 %
Level 4	526	25.7 %	25.7 %
Grades 6–8			
Level 1	—	15.9 %	100.0 %
Level 2	425	28.5 %	84.1%
Level 3	489	25.9 %	55.6 %
Level 4	534	29.8 %	29.8 %
Grade 10			
Level 1	—	16.1 %	100.0 %
Level 2	434	30.1 %	84.0 %
Level 3	498	28.9 %	53.9 %
Level 4	541	24.9 %	24.9 %

Exhibit 15: Percentage of Students at Each Performance Level—Science

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	19.8 %	100.0 %
Level 2	430	18.2 %	80.2 %
Level 3	469	17.5 %	62.0 %
Level 4	496	44.5 %	44.5 %
Grades 6–8			
Level 1	—	22.1 %	100.0 %
Level 2	447	18.5 %	77.9 %
Level 3	489	15.3 %	59.3 %
Level 4	514	44.0 %	44.0 %
Grade 10			
Level 1	—	25.3 %	100.0 %
Level 2	463	25.0 %	74.7 %
Level 3	506	16.1 %	49.7 %
Level 4	535	33.6 %	33.6 %

Exhibit 16: Percentage of Students at Each Performance Level—Social Studies

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	19.3 %	100.0 %
Level 2	423	32.7 %	80.7 %
Level 3	492	30.1 %	48.1 %
Level 4	549	18.0 %	18.0 %
Grades 6–8			
Level 1	—	19.7 %	100.0 %
Level 2	439	27.3 %	80.3 %
Level 3	503	34.1 %	53.0 %
Level 4	560	19.0 %	19.0 %

Chapter 5: Technical Characteristics and Interpretation of Student Scores

This section describes the psychometric analyses conducted as part of the South Carolina Alternate Assessment (SC-Alt) 2007 and 2008 operational administrations. Only the social studies assessment was calibrated and evaluated using 2008 operational data. However, in order to provide a complete description of the technical characteristic of the 2008 assessment in all content areas, this chapter also reports the data analysis results for the English language arts (ELA), mathematics and science sections of the assessment which had previously been calibrated using the 2007 operational data (see American Institutes for Research and South Carolina Department of Education, 2008, *South Carolina's Alternate Assessment, SC-Alt Spring 2007 Operational Administration*). The reported analyses are intended to ensure the quality of the items, the assessment materials and instruments, and the score reporting scales as measures of state academic standards.

As a reminder to the reader, there are three grade-band forms in each content area: elementary school (grades 3–5), middle school (grades 6–8), and high school (grade 10; excepting social studies). At each grade band, the assessments have three potential starting tasks that correspond to three levels of task complexity (high, moderate, and low). Students are assigned to a starting task on the basis of teacher judgments recorded in the Student Placement Questionnaire (SPQ) for each content area. Linking tasks connect the grade-band forms so that the vertical test scale could be created.

Analysis and Scaling of Items, Tasks, and Test Forms

The ELA, mathematics, science, and social studies assessments received comprehensive psychometric analyses, including initial item calibrations, after their earlier field-testing. Final calibrations were estimated for the ELA, mathematics, and science content areas on the basis of operational data gathered during the spring 2007 operational administration; final calibrations for social studies were computed from operational data of the spring 2008 administration. Calibrations based on operational data were considered superior to those based on field-test data. The vertical scales were also defined using the linking tasks as the vehicle that connected the elementary, middle, and high school forms.

AIR calibrated the operational items, estimated examinee proficiencies, and calculated scale scores and achievement levels for operational forms. This process entailed examining item statistics to ensure quality measurement across the range of the assessment, calibrating the items within each content area to a common scale, then applying a maximum-likelihood (ML) scoring algorithm to each student's responses to estimate his or her proficiency scores and assign the correct achievement level.

Assignment of Examinees to Starting Tasks and Item Calibration and Test Forms Linking

All eligible students participated in the spring 2008 test administrations. The case counts of somewhat over 1,100 students each in elementary and middle school and approximately 350 students in high school, per content area, enabled effective calibration across task starting points and grade bands. Students were assigned to one of three starting points on the basis of the sum of the teacher responses on the SPQ. The SPQ cut scores were shown to correlate with student

achievement scores on the 2006 field-test administrations (for details, see American Institutes for Research, 2008, *South Carolina Alternate Assessment (SC-Alt): Technical report for English Language Arts and Mathematics Field Test Administration, Spring 2006*). The assignment of student starting tasks based on the SPQ cut scores was intended to expose students to items that were ideally suited to their current level of achievement while ensuring that (a) each student responded to an adequate number of items so that reliable and content-valid proficiency scores could be estimated and (b) an adequate number of students responded to each item for the joint calibration to be reliable.

Teachers were instructed to administer all tasks associated with the assigned starting point, with provisions for dropping to a lower starting point (task) if the student was unable to respond to the items in the task at the assigned starting point. Students who were assigned to high and moderate levels of the assessment but were unable to respond to items in the tasks at those levels may have been moved back to a less difficult starting point.

The linking design allowed a joint (concurrent) calibration of all items within a content area and the placement of the items on a common difficulty scale. The tasks actually used to link the grade-band forms (linking tasks) were selected, in part, on the basis of their moderate difficulty levels. Moderately difficult tasks contribute to more stable linking across levels than tasks that may be either too easy or too difficult for the examinees.

Linking across grade-band forms was accomplished by using common tasks across grade bands. Some of the tasks from the elementary form are on the middle school form; some of the tasks from the middle school form are on the high school form. (For details, see "Linking Tasks in Each Grade-Band Assessment" above.) In general, tasks are assigned in such a manner that the forms increase slightly in difficulty as examinees progress through the grade bands. This means that a task assigned to the moderate level of complexity in the elementary form may be assigned to the low-moderate level in the middle school form.

A similar linking design was employed for the social studies field-test forms.

See Appendix E for a summary of the linking design in each of the four content areas.

Analysis Plan

AIR's analyses presented in the remainder of this chapter were conducted in five steps:

1. data preparation and quality control,
2. classical item analysis,
3. review of items not meeting psychometric criteria for inclusion on operational forms,
4. joint calibration of items according to the Rasch model, and
5. final achievement estimation and scale score calculation for operational forms.

Data Preparation and Quality Control

Before analyzing the operational test data, AIR psychometricians performed a number of quality control procedures to ensure that scanning operations resulted in accurate data capture of

the teacher-recorded student responses. Prior to the test administration, AIR verified all of the point values for each form's answer folder. For each form, two AIR staff members independently verified the possible responses and point values for each item.

After receiving the scanned test data, AIR analysts carefully examined the data file to verify its accuracy. Descriptive statistics were computed to ensure that student case counts on the pre-identification file generally corresponded to the actual counts based on test data at the state, school, and classroom levels. In addition, AIR verified that the total number of items in the data file matched the number of items on the answer folder and in the test booklet and then examined the frequency distributions of item responses to identify potential scoring problems, such as out-of-range values or unused response categories.

For purposes of item analysis and student scoring, respectively, non-response (NR) data were treated in two different ways:

For *item analysis and calibration purposes*, a student had to have at least three scored responses for the testing attempt to be considered valid. For a response to be considered a scored response, the test administrator had to have assigned a numeric score (0–4) to the student's response. If the administrator scored NR for all items in a task, the task was treated as not administered, and NR values were recoded as missing.

For *operational scoring* of student responses and estimation of student proficiency, however, the NR codes were treated as indications that the item was administered and that the student did not possess the content area knowledge and skill to respond. In this case, all NR values were recoded as zeroes and included in the student proficiency estimates. Following this recoding, tests were reexamined to determine the number of scored responses (0–4) in each content area. For operational scoring, a student had to have at least five scored responses of any kind for the assessment to be considered a valid attempt within a content area.

After the accuracy of the data file was verified, classical item analyses and Item Response Theory (IRT) analyses were performed. Several quality control procedures were taken to ensure the accuracy of these analyses.

As an initial step, the program control file was checked by two data analysts to ensure that form layout was correctly specified and that item response values were correct. As a second step, two analysts independently performed all analyses. Results of the parallel analyses were compared for mistakes by using commercially available file comparison software. Last, the analysis results were spot-checked by using other commercially available statistical software to ensure that the results were consistent across statistical software packages. *These comprehensive quality control steps are highly effective in detecting any issues that might influence the interpretation of the item analysis results.*

Classical Item Analysis

Classical item analysis for the SC-Alt operational and field-test forms was conducted using the *AM* statistical software (http://am.air.org). The item analysis yielded the proportion of

students in each response category, the percentage of omitted responses for that item,[3] and the proportion of students who were unable to respond to the item because of access limitations (where relevant). Correlations between the item score and the test score were computed using adjusted polyserial correlations. For purposes of calculating item statistics, omitted items were treated as incorrect when there was at least one scored response within the same task (see above). Minimum and maximum point values, average item scores, and adjusted item-total polyserial correlations were calculated for all items.

Test form statistics, such as internal consistency reliability estimates and standard error of measurement statistics, were suppressed at this point because all students were not expected to take all items. Such statistics would be misleading before Rasch scoring was applied. Special marginal reliability analyses used to determine the reliability of the student score estimates are described in a later section of this chapter.

The proportion of students in each score-point category was calculated as defined by the item's scoring guidelines, as well as the proportion of students with blank responses within attempted tasks (i.e., those with at least one scored response). Item difficulty was computed as the mean score on the item across all students taking the form and with a scored response on that item. The average proportion of total points, calculated as the mean score divided by the total number of points possible on the item, serves as an additional measure of item difficulty.

Review of Items Not Meeting the Specified Psychometric Criteria

Classical item analysis provided information about the technical quality of the items; items failing to meet specified psychometric criteria were flagged for subsequent review. During field-testing of ELA and mathematics (spring 2006), science (fall 2006), and social studies (spring 2007), AIR reviewed all flagged items in concert with SCDE to determine whether they were of sufficient psychometric quality. For the 2007 operational forms in ELA, mathematics, and science, and for the 2008 operational form in social studies, AIR conducted a statistical review of the items to determine whether any operational items were performing in an unacceptable fashion.

All item reviews were conducted using polytomous item flagging rules that AIR has put in place for other alternate assessments. Items were flagged for review for any of six reasons:

- Item-total score correlation was negative.
- The mean proportion correct did not decrease for each successive score category.
- Item difficulty value indicated that the item was extremely difficult or extremely easy.
- The omit rate was high.
- Differential item functioning (DIF) was present.
- Item fit statistics were unacceptable.

[3] An item was considered omitted if no response was recorded for the item (or the test administrator marked NR on the student score sheet), but the student responded to subsequent items on the task.

Adjusted polyserial correlations were flagged if they were less than .20. Near zero or negative adjusted polyserial correlations may indicate a flawed scoring rubric, mis-ordering of response categories, reader difficulties in interpreting the rubric, or an item that does not measure the construct of interest.

Items were also flagged if the mean proportion correct (of items attempted) of students in a score-point category was lower than the mean proportion correct of students in the next lower score-point category. For example, an item was flagged if, on average, students who received 3 points on the item got a lower proportion of the total points possible on items with scored responses than students who received only 2 points on the item. This situation may indicate that the scoring guidelines are flawed. This flag was interpreted conservatively because students may take items with different average difficulty levels.

Items were flagged if the proportion of students in any score-point category was greater than .95. A very high proportion of students in any single score-point category may suggest that the other score points are not useful or, if the score point is the minimum or maximum score-point category, that the item may not be appropriate for students at that complexity level. Analysts must take into account the item-total score correlation as well as possible instructional factors when interpreting the statistic.

Items with omit rates greater than 5% were flagged. High rates of response omission may indicate confusion by test takers or administrators on how to respond to the item, excessive test speededness, or an item that was too difficult. It was expected that rates of omission may be somewhat higher for the alternate assessment population, so this was not considered a critical problem.

AIR conducted analyses of differential item functioning (DIF) on all test items to detect potential item bias. The purpose of these analyses was to identify items that may favor students in one group over students of similar achievement in another group. AIR conducted DIF analyses that compared African American/White and female/male student subgroups. The sample sizes were very small for each subgroup (see the demographic tables in the previous section); thus, DIF analyses have limited utility. However, DIF statistics were calculated for the purposes of item review.

For polytomous items, both the Mantel-Haenszel chi-square (*MH* χ^2; Zwick & Thayer, 1996; Zwick, Donoghue, & Grima, 1993) and the Standardized Mean Difference (SMD; Dorans & Kulick 1986) were calculated. The classification rules are defined in Exhibit 17. Items in the "C" DIF category, indicating evidence of differential item functioning on the items, were flagged for review.

Exhibit 17: Summary of DIF Classification Rules for Polytomous Items

DIF Category	Rule
C	The p-value of MHχ^2 is less than .05 and |SMD|/|SD| is greater than 0.25.
B	The p-value of MHχ^2 is less than .05 and |SMD|/|SD| is greater than

	0.17 and less than 0.25.
A	The p-value of $MH\chi^2$ is not significant at the .05 level or |SMD|/|SD| is less than 0.17.

In addition, items were flagged on the basis of criteria set for INFIT and OUTFIT statistics produced by Winsteps. To evaluate item fit, we examined the item fit statistics provided by Winsteps. The mean square INFIT and mean square OUTFIT statistics reported by Winsteps are based on weighted and unweighted standardized residuals for each item response, respectively. These statistics indicate the discrepancy for each item between observed item responses and the item responses predicted under the Rasch model. Both fit statistics have an expected value of 1. Values substantially greater than 1 indicate unmodeled noise (model underfit), and values less than 1 indicate a lack of stochasticity (model overfit).

Because it is weighted by the variance of a person's response to an item, the INFIT statistic is sensitive to deviations from expected response patterns among high information items (i.e., items with locations near the theta estimate for the student) that could indicate structural problems with test items or test form construction (e.g., mis-keyed items, items not measuring the common underlying construct; conversely, excessive redundancy in item content resulting in overdetermined response patterns).

The OUTFIT statistic, an unweighted mean square, is sensitive to low information responses (e.g., easy items missed by high-ability students, difficult items correctly answered by low-ability students) and may therefore indicate the presence of outliers (Linacre & Wright, 1994). Items were flagged if the mean square INFIT or OUTFIT values were less than 0.7 or greater than 1.3. Misfitting items were evaluated in conjunction with the classical item analysis results to determine whether items should be included in the operational pool. We reviewed item fit as part of the scaling process and the item-data review process.

Items flagged on the basis of any of the aforementioned criteria were reviewed by AIR psychometricians and SCDE officials. First, a team of AIR psychometricians reviewed all flagged items to ensure that the data were accurate and properly analyzed, that response keys were correct, and that there were no obvious problems with the items. AIR recommended whether the item should be retained in the item pool or discarded, depending on the reason for the item flag and its effect on the quality of the assessment as a whole. SCDE had the final authority on whether the flagged items should be included in the operational scoring based on the item statistics and content appropriateness of the items. Additional information on the SCDE review of operational item statistics is available from SCDE in the Response of SCDE to the South Carolina Educational Oversight Committee Recommendations, November 19, 2007.

No items on the operational forms (ELA, mathematics, science) were found to violate the psychometric criteria so severely that they needed to be removed from scoring.

Item Response Theory Calibration and Linking Test Forms

This section describes AIR's procedures for item calibration using IRT techniques. Item parameters were estimated using the Partial Credit model (Masters, 1982) approach available using Winsteps software. A common item design was used to enable simultaneous calibration

and linking across grade-band test forms in each content area. Items were jointly calibrated across grade bands in a single Winsteps run for each content area. This calibration approach put the item parameters of all grade-band test forms within a content area on the same scale.

For 2008 the results reported on the vertical scale are in Exhibit 18 and Exhibit 19. Several things to notice in these exhibits are that the mean scores show a general upward trend across grades. This indicates that a vertical scale is a useful way to describe the results of this population of students. Second, in almost every grade band, a few students were at the floor of the test (minimum scaled score equal to 260), but very few reached the ceiling (maximum scaled score equal to 740). The exceptions are the mathematics and science performance in high school, which are essentially equal to the middle school results.

Exhibit 18: Scale Score Statistics, by Grade Band, Overall

	Elementary					Middle					High				
Subject	**N**	**Mean**	**SD**	**Min**	**Max**	**N**	**Mean**	**SD**	**Min**	**Max**	**N**	**Mean**	**SD**	**Min**	**Max**
ELA	1231	487	75	260	704	1048	504	81	260	707	373	512	87	260	740
Math	1228	489	74	260	698	1043	503	82	260	705	372	501	81	260	718
Science	881	487	69	260	733	759	503	80	260	729	370	504	90	260	740
Social Studies	847	495	79	260	740	723	511	90	260	732	0	.	.	.	.

Exhibit 19: Scale Score Statistics, by Grade Band, by Primary Disability

	Elementary					Middle					High				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
ELA															
Overall	1231	487	75	260	704	1048	504	81	260	707	373	512	87	260	740
Severe Mental Disability	128	388	69	260	503	111	394	63	260	526	35	398	90	260	688
Moderate Mental Disability	371	474	44	260	595	439	496	58	260	707	181	510	69	260	729
Mild Mental Disability	327	545	52	426	704	246	569	64	260	707	66	574	66	429	729
Autism	252	476	61	260	704	153	487	66	260	707	45	490	79	260	671
Math															
Overall	1228	489	74	260	698	1043	503	82	260	705	372	501	81	260	718
Severe Mental Disability	128	393	64	260	494	108	393	62	260	513	35	390	84	260	545
Moderate Mental Disability	371	475	45	260	612	439	490	59	260	650	181	497	62	260	718
Mild Mental Disability	325	546	57	393	698	245	573	66	387	705	66	564	64	441	718
Autism	252	483	56	260	698	152	489	66	260	650	45	489	75	324	668
Science															
Overall	881	487	69	260	733	759	503	80	260	729	370	504	90	260	740
Severe Mental Disability	93	391	66	260	512	81	387	67	260	506	35	372	87	260	577
Moderate Mental Disability	264	480	48	260	616	307	496	65	260	729	180	506	69	260	740
Mild Mental Disability	225	535	41	396	733	181	566	51	394	729	65	569	67	410	740
Autism	189	481	54	260	656	118	482	65	260	631	45	484	89	260	740
Social Studies															
Overall	847	495	79	260	740	723	511	90	260	732	0	.	.	.	.
Severe Mental Disability	93	384	78	260	513	75	379	74	260	544	0	.	.	.	.
Moderate Mental Disability	246	484	53	260	630	306	506	62	260	667	0	.	.	.	.
Mild Mental Disability	234	553	49	452	740	162	590	69	344	732	0	.	.	.	.
Autism	170	489	61	312	740	111	492	73	260	667	0	.	.	.	.

An important feature of the South Carolina alternate assessment is the vertical scale that permits the measurement of student progress on the state content standards over time. Such a scale provides educators and parents with information that can be useful for monitoring student performance as students move through grades over time.

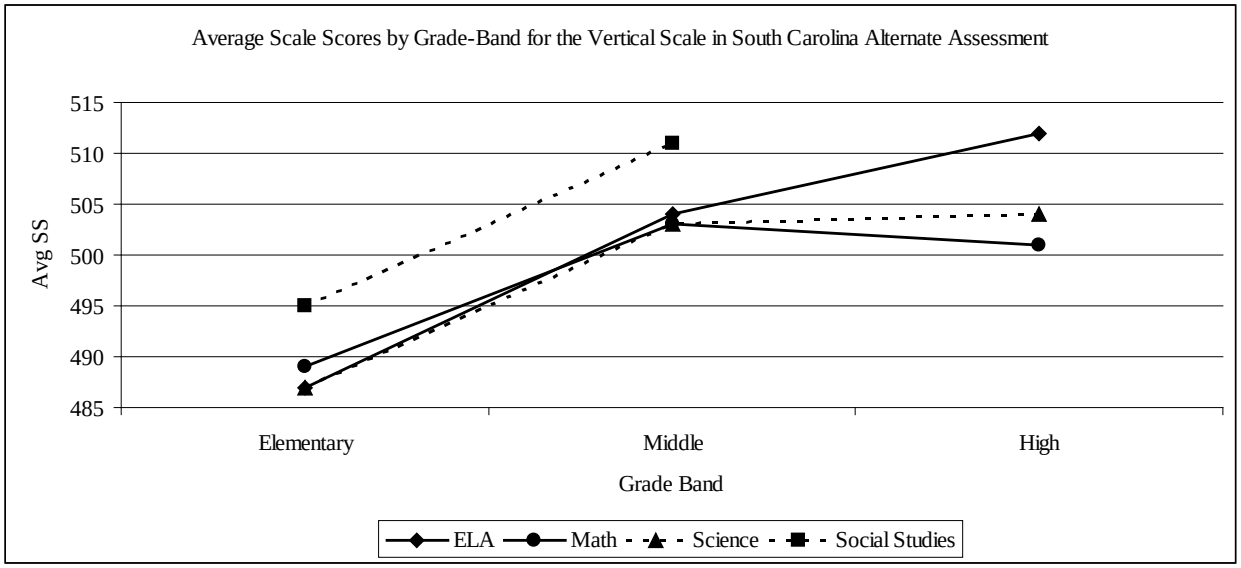
The development of this scale required the use of a common item linking design. In a common item design, *linking items* appear on the higher grade-band assessment as well as on the lower grade-band assessment. These linking items allow for the grade-band scales to be connected, thus establishing the vertical scale.

There are at least two features of this linking design that warrant clarification. First, the linking items are the same (i.e., they are the same items) between two grade-band forms. They are connected to grade-specific standards in the higher grade as well as in the lower grade. As a result, students are never exposed to off-grade-level content since the common items serve a dual purpose in measuring content in both grade bands. Second, even though the same forms will be used the first two years of administration (pending planned new task and item development), it is not likely that exposure of the items to the students would trigger responses based on the recollection of any item's administration the previous year. As students grow academically, their starting task will likely change each year. New starting tasks mean that a portion of the items any student receives will be unique each year.

The South Carolina Department of Education is in the process of evaluating growth models under a federal grant (Modifications, Accommodations, Reports, and Standards). The use of the extant vertical scale for SC-Alt will facilitate deliberations for the possible implementation of a growth model for the alternate assessment.

A graph of the overall pattern of performance on the vertical scale is shown in Exhibit 20. Again, there is a general upward trend across all grade bands in each of the four subject areas. This graph shows that the vertical scale in the South Carolina Alternate Assessment was successful at capturing growth across grade bands.

Exhibit 20: Overall Pattern of Performance on the Vertical Scale



Using Item Responses to Estimate Student Proficiency

This section describes the estimation of student proficiency for the SC-Alt operational administration of English language arts, mathematics, and science assessments for elementary, middle, and high school; social studies assessments for elementary and middle school are also reported. The section first describes the estimation procedures used to determine student proficiency based on the items administered, then the transformation of proficiency estimates on the Rasch theta scale into scale scores, and finally the relation of achievement estimation to reliability estimation.

Student proficiency scores were estimated using a maximum-likelihood approach based on the scored items for each student.[4] This method calculates the theta score that maximizes the likelihood function of the given item responses for each student. Comparable scale-score estimates from these different item responses were achieved through the measurement-invariance property of IRT ability estimates, even when students were exposed to different ranges of items.

Under the Rasch-based IRT model, there is a one-to-one correspondence between the estimated theta score and the total raw score for a specific set of items. However, in the SC-Alt assessments, each student can take different sets of items. Using the pattern scoring method for calculating theta scores, we ensured that (a) two students who took the same items and achieved the same item scores were assigned the same theta score, and (b) students who took more difficult items were assigned higher theta scores than students with the same raw scores who took less difficult items. Thus, the scoring method took into account both the number of raw score points the student achieved and the difficulties of the items the student responded to. This scoring process was performed separately for each content area.

[4] The first step in this process was to rescore student responses consistent with the operational scoring method described under the "Data Preparation and Quality Check" section.

Once theta values had been estimated for each student, AIR converted the theta estimates to scale scores using a scale metric determined by SCDE in consultation with AIR. The SC-Alt assessments were scaled to have a mean of 500 and a standard deviation of 80 on the vertical scale for the grade band 6–8 assessment. The grade-band 3–5 and grade 10 assessment means and standard deviations were calculated in relation to grade-band 6–8 mean and standard deviation. This was done by performing a linear transformation of the Rasch theta scale for each content area, fixing the mean of the middle school test form scale at 500, and multiplying the student's theta deviation score by 80 as shown in the formula below:

$$y_{ijk}* = 500 + (\frac{\hat{\theta}_{ijk} - \hat{\mu}_k}{\hat{\sigma}_k}) * 80,$$

where

i indexes student;

j indexes grade band;

k indexes content area;

y^*_{ijk} is the scale score for student *i* in grade band *j* and content area *k*, given estimated ability, θ_{ijk};

$\hat{\mu}_k$ is the content-area-specific mean for the middle school test form; and

$\hat{\sigma}_k$ is the content-area-specific standard deviation for the middle school test form.

SCDE also decided to truncate the scale score ranges so that the lowest possible scale score was 260 and the highest possible scale score was 740. Student scale score estimates were truncated to the smallest whole integer (e.g., an estimated scale score of 440.60 would become 440). Additionally, scale scores were calculated and checked using a method similar to the process for total raw data.

Once scoring was completed, it was possible to estimate the internal consistency score reliability of the grade-band assessments by estimating the marginal measurement error across students. These estimates produced different standard errors for each student, depending on the items they were given and their level of performance on those items. Then, this value was used to determine the score reliability as the proportion of true score variance to observed score variance. We estimated this value within each content area (a) across the entire theta scale, (b) across grade-band forms, and (c) for each starting point within a grade band.

Test Score Reliability

This section provides the marginal reliability for each grade band, content area, and groups of students beginning at each starting task determined by the SPQ for the spring 2008 administration.

Classical test theory-based reliability indices, such as Cronbach's alpha, were not appropriate for the SC-Alt because the length of the test and the subset of items differed for each student. The reliability coefficient for the SC-Alt was, therefore, calculated as the *marginal reliability* (Sireci, Thissen, & Wainer, 1991), which is equivalent in interpretation to classical internal consistency estimates of reliability.

First we determined the marginal measurement error variance, $\overline{\sigma}_{e^*}^2$, across all examinees:

$$\overline{\sigma}_{e^*}^2 = \int \sigma_{e^*}^2 p(\theta) d\theta = \frac{\sum \sigma_{e^*}^2}{N},$$

where $\sigma_{e^*}^2$ is the square of the standard error of student ability estimate, $\hat{\theta}$. Thus, the marginal measurement error variance could be estimated as the average of squared standard error of $\hat{\theta}$.

Then we estimated the marginal reliability as

$$\overline{\rho} = \frac{\hat{\sigma}_{\theta}^2 - \overline{\sigma}_{e^*}^2}{\hat{\sigma}_{\theta}^2},$$

where $\hat{\sigma}_{\theta}^2$ is the variance of observed θ estimates.

The marginal reliability estimate, $\overline{\rho}$, can be interpreted similarly to classical reliability indices such as Cronbach's alpha. Estimates of the marginal reliability for the test forms corresponding to the three SC-Alt grade-band assessments can be seen in Exhibit 21.

All marginal reliability estimates at each grade band for ELA and mathematics assessments exceeded 0.90. The science assessment also exceeded a reliability of 0.90 on the elementary and middle school forms, with the only exception of 0.882 for the high school form (Grade 10). The marginal reliability estimates for the social studies assessments were slightly lower, 0.891 and 0.865 at the elementary and middle school grade bands, respectively. The reliability estimates of all four content areas fall into the range of reliability coefficients found with large-scale assessments (Rudner and Schafer, 2001) and, subject to NCLB assessment system review, meet the reliability requirements for assessments used for the purposes for which the SC-Alt was designed.

In addition to the marginal reliability estimates, Exhibit 21 also displays the marginal standard errors of measurement for each subject and grade band, labeled $\overline{\sigma}_{e^*}$. These marginal standard errors of measurement range between 23 and 33 scale score units, placing the SEM at approximately one third of a standard deviation of the content area and grade band.

Exhibit 21: Marginal Reliability and Standard Error of Measurement by Grade Band and Subject

Subject	Grade Band	Elementary	Middle	High
English Language Arts	N	1229	1050	373
	Reliability	0.909	0.908	0.907
	$\overline{\sigma}_{e^*}$	22.7	24.5	26.6
Math	N	1226	1045	372
	Reliability	0.903	0.906	0.906
	$\overline{\sigma}_{e^*}$	23.0	25.4	24.9
Science	N	879	761	370
	Reliability	0.901	0.909	0.882
	$\overline{\sigma}_{e^*}$	21.6	24.4	31.0
Social Studies	N	846	724	x
	Reliability	0.891	0.865	x
	$\overline{\sigma}_{e^*}$	26.0	33.1	x

Appendix G shows the marginal reliability estimates broken out further by groups of students beginning at each starting task, as determined by the SPQ. The reliability coefficients in Appendix G are generally somewhat attenuated compared with those in Exhibit 21, due to the reduction in variance of scale scores grouped by starting task.

Classification Accuracy

This section describes the extent to which student achievement-level classifications were accurate across students. Classification accuracy was estimated for each cut score as the average probability of correct performance-level assignments across all examinees (assignments above or below the cut score), given each examinee's estimated proficiency score, θ_i:

$$\mathrm{CA_K} = \frac{\sum_{k \geq K} P(\theta_i > \theta_K^* \mid \theta_i, k_i) + \sum_{k<K} 1 - P(\theta_i > \theta_K^* \mid \theta_i, k_i)}{N},$$

where

θ_i is the proficiency (i.e., theta) of student *i*;

k_i is the performance level of student *i*;

θ_K^* is the cut score for the performance level *K* on the theta scale; and

N is the sum of the number of students at or above the cut score, $N_{k \geq K}$, and the number of students below the cut score, $N_{k<K}$, or simply the total number of students.

Thus, $P(\theta_i > \theta_K^* \mid \theta_i, k_i)$ is the probability of a student with θ_i and the performance level k_i to be above the cut score *K*. The classification accuracy is the expected rate of correct classification probability; therefore, the higher value indicates the superior classification accuracy. Exhibit 22 shows the classification accuracy by content areas, performance levels, and grade bands.

Exhibit 22: Classification Accuracy

Subject	Performance Level Cut Score	Grade Band			Overall
		3–5	6–8	10	
English Language Arts	Level 2	0.927	0.928	0.929	0.927
	Level 3	0.846	0.899	0.889	0.873
	Level 4	0.889	0.908	0.898	0.898
Mathematics	Level 2	0.893	0.882	0.889	0.888
	Level 3	0.876	0.901	0.866	0.884
	Level 4	0.868	0.887	0.890	0.879
Science	Level 2	0.875	0.901	0.870	0.884
	Level 3	0.876	0.896	0.837	0.877
	Level 4	0.863	0.883	0.852	0.870
Social Studies	Level 2	0.911	0.910	x	0.911
	Level 3	0.869	0.879	x	0.874
	Level 4	0.803	0.803	x	0.803

For example, according to the estimates in Exhibit 22, 93% of students were correctly classified (a) in Level 1 or (b) in Level 2 or above for the grade band 3–5 English language arts assessment. All students in all grade bands had a probability greater than .846 of being classified accurately as proficient or higher (i.e., as level 3 or 4).

The calculation of probability of the correct performance level for students is described in the following section.

Calculation of the Probability of Being Classified Above a Cut Score Given the Student's Theta Score

For each student we can compute the likelihood of theta $L(\theta \mid \mathbf{z}, \mathbf{b})$. Suppose that the prior of the theta distribution is $f(\theta)$. Then, using Bayes' rule, we have

$$f(\theta \mid \mathbf{z}, \mathbf{b}) \propto f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}),$$

where $L(\theta \mid \mathbf{z}, \mathbf{b})$ is the likelihood of theta given the response $\mathbf{z}$ and item parameters $\mathbf{b}$; hence, the probability at and above cut is

$$P = \frac{\int_{\theta \geq \theta_{\text{cut}}} f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}) d\theta}{\int f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}) d\theta} ,$$

where $f(\theta)$ can take different distribution such as normal, or uniform, depending on our prior belief.

Calculation of $L(\theta \mid \mathbf{z}, \mathbf{b})$

For the Rasch model, we have

$$L(\theta \mid \mathbf{z}, \mathbf{b}) = \prod_{i \in MC} \left(\frac{Exp(z_i \theta - b_i z_i)}{1 + Exp(\theta - b_i)} \right) \prod_{i \in CR} \left(\frac{Exp(z_i \theta - \sum_{k=1}^{z_i} b_k)}{1 + \sum_{i=1}^{K_i} Exp(\sum_{k=1}^{i} (\theta - b_k))} \right)$$

$$\propto Exp(r\,\theta) \prod_{i \in MC} \left(\frac{1}{1 + Exp(\theta - b_i)} \right) \prod_{i \in CR} \left(\frac{1}{1 + \sum_{i=1}^{K_i} Exp(\sum_{k=1}^{i} (\theta - b_k))} \right)$$

where K_i is the maximum score for item *i* when this item is a CR item. It can be noted that the calculation above depends on total raw score r only when using the attempted items.

Chapter 6: Score Reports

This chapter describes the method used for reporting scores on the SC-Alt for the spring 2008 administration. An Individual Score Report (ISR) is included in Appendix H as an example of the highly detailed and diagnostic nature of the reports. This chapter gives a brief overview of how scores on the SC-Alt assessments are reported; a more detailed description is available in a separate *Score Reports User's Guide*.

The SC-Alt assessments have three types of score reports: the ISR, or family report; school reports; and district reports. Each report conveys specific information to its target audience. The reports are designed to be easily used by parents and educators. Of particular note, the reports include in-depth information about what students know and can do relative to the South Carolina academic content standards and to the performance levels.

The ISR provides specific performance feedback for each student across four content areas: mathematics, English language arts, science, and social studies (added in 2008). Within each content area, a graphic bar highlights the student's performance level along the proficiency scale. Each performance level is described in broad, easy-to-understand content terms. Further descriptions of what a student knows and can do are tailored and printed for each obtained performance level. For example, if a student is classified as Level 3 in mathematics, the following message is printed: "Students who score at Level 3 should be able to add and subtract simple numbers, count and compare objects in a group, compare objects by color, size, or shape, identify three-dimensional shapes, and read information in a graph." (Note: Scale scores were added to the ISR starting with the spring 2008 reporting cycle.)

Specific activities, based on each student's performance level for each content area, are presented for the family to do at home to help ensure positive academic growth in the content area.

The school report provides a summary of the performance of each student in the school. The alphabetical list of students contains basic demographic information and test form administered, in addition to achievement data. A scale score and achievement level are listed for each student for each content area. A school summary shows the number of students scoring at each performance level.

Three district-level reports are issued. The district roster summary report displays the roster of the district's tested students along with their demographic information, their scale scores and performance levels for each content area, and type of test form. The district summary by test form report presents a roster of schools in which students were tested, identifying the test form and giving the number of students tested in each content area and the percentages achieved in each performance level by content area. The total number of students tested with each form and their performance-level distributions by content area are listed at the bottom of the report. The district demographic summary report shows the number of students tested and the distributions of performance levels in all content areas, disaggregated by gender, ethnicity, lunch program, migrant status, and ESL status.

The separate *Score Reports User's Guide* has more specific information on how to interpret student scores and score reports and how to relate academic growth as measured by the SC-Alt to classroom curricula and activities. The guide has been widely distributed throughout South Carolina.

Chapter 7: Student Performance Data from the Spring 2008 Administration

Performance data from the spring 2008 administration are presented in this chapter. This was the second operational administration of the SC-Alt ELA, Mathematics and Science assessments and the first operational administration of the SC-Alt Social Studies assessment.

A total of 2,657 students[5] from 84 school districts and 476 schools were tested with the SC-Alt in spring 2008. The total number of tested students with one or more valid content area scores was 1,232 for the elementary form, 1,051 for the middle school form, and 373 for the high school form. Five students tested on two forms each; as it happened in each case, social studies was tested on a different grade band than the other content areas.

One-half of the participating school districts (42; 50%) tested 15 or fewer students; 28 districts (33%) tested 15 to 50 students; and 14 districts (17%) tested 51 to 228 students each. Only five districts tested more than 100 students.

Of the 476 schools testing SC-Alt students, 300 (63%) tested five or fewer students; 119 (25%) tested six to 10 students; 46 (10%) tested 11 to 20 students; and 11 schools (2%) tested 21 to 56 students. Only two schools tested more than 50 students each (54 and 56 students).

The elementary school form was developed to be administered to students who are 8, 9, or 10 years old at the beginning of the school year, which are the ages typical of students enrolled in grades 3, 4, and 5. The middle school form was developed for students of ages 11, 12, and 13 (typical of students enrolled in grades 6, 7, and 8), and the high school form was developed for students age 15 (typical age of students in grade 10).

Students who are reported as tested with the elementary and middle school forms with ages outside the specified ages were assigned the forms by the test administrator by error or, in some cases, have errors in the coding of their birth dates on the data files. The numbers of students reported outside the expected ages for the elementary and middle school forms is approximately 1% for each content area. Students reported as having been tested on the high school form with ages below 15 appear due to form assignment or birth date errors. Students older than 15 (e.g., 16) are assessed with the high school form if they have not been assessed at the high school level previously.

The performance of students by grade-band form, age, and demographic group for the ELA, mathematics, science, and social studies content areas is presented in Exhibit 23 to Exhibit 26.

[5] One of the students passed away and was excluded from reporting; the effective total population size is 2,656.

Exhibit 23: Performance by Grade-Band Form and Student Age—ELA

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	389	473	72	42	58	.	.	.	.	.	.	.	.	.	.
9	427	490	72	33	67	.	.	.	.	.	.	.	.	.	.
10	402	497	79	30	70	1	423	.	100	.	.	.	.	.	.
11	7	562	53	.	100	355	495	75	39	61	.	.	.	.	.
12	2	350	127	100	.	339	506	76	32	68	.	.	.	.	.
13	2	474	30	50	50	353	512	89	31	69	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	1	548	.	.	100
15	.	.	.	.	.	.	.	.	.	.	370	512	88	36	64
Blank	1	503	.	.	100	.	.	.	.	.	.	.	.	.	.
STUDENT'S ETHNICITY															
African American	647	492	78	32	68	536	513	83	30	70	205	514	93	32	68
White/Asian	5	486	59	40	60	.	.	.	.	.	.	.	.	.	.
Other	3	501	86	33	67	4	497	30	25	75	.	.	.	.	.
African American/American Indian	3	505	39	33	67	.	.	.	.	.	1	583	.	.	100
American Indian	5	492	146	40	60	3	509	86	33	67	.	.	.	.	.
Asian	8	493	42	25	75	11	475	46	55	45	2	521	9	.	100
Hawaiian/Pacific Islander	1	404	.	100	.	1	484	.	.	100	1	508	.	.	100
Hispanic	47	482	66	40	60	35	480	75	37	63	3	541	54	.	100
White	494	480	72	37	63	445	497	78	38	62	158	509	82	42	58
White/African American	17	491	90	35	65	11	502	86	36	64	1	462	.	100	.
White/American Indian	.	.	.	.	.	2	507	72	50	50	.	.	.	.	.
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
STUDENT'S GENDER															
Female	405	485	82	36	64	355	505	86	33	67	137	509	87	34	66
Male	825	488	72	34	66	693	504	78	35	65	234	515	88	36	64

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															
Pre-functional	31	470	75	45	55	22	477	64	36	64	2	515	17	.	100
Beginner	1	537	.	.	100	1	520	.	.	100	.	.	.	.	.
Intermediate	.	.	.	.	.	2	525	57	.	100	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	1	430	.	100	.	1	501	.	.	100	.	.	.	.	.
Title III Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English speaker I	.	.	.	.	.	1	475	.	100	.	.	.	.	.	.
English speaker II	1191	487	76	35	65	1018	505	81	34	66	364	512	88	35	65
Pre-functional Waiver	1	473	.	.	100	.	.	.	.	.	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	1	598	.	.	100	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	5	469	50	60	40	2	510	49	50	50	5	509	98	60	40
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	386	476	75	39	61	333	490	73	41	59	113	493	83	45	55
Free	745	494	76	32	68	627	512	83	30	70	225	522	88	30	70
Reduced	97	476	64	40	60	88	500	87	33	67	30	521	90	33	67
Blank	2	459	58	50	50	.	.	.	.	.	3	443	45	100	.
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)															
Severe Mental Disability	128	388	69	90	10	111	394	63	94	6	35	398	90	91	9
Moderate Mental Disability	371	474	44	39	61	439	496	58	34	66	181	510	69	34	66
Mild Mental Disability	327	545	52	6	94	246	569	64	5	95	66	574	66	9	91
Autism	252	476	61	41	59	153	487	66	42	58	45	490	79	42	58
Deaf/Blindness	3	360	173	67	33	3	512	48	.	100	4	537	57	25	75
Emotional Disability	7	572	75	.	100	4	618	61	.	100	.	.	.	.	.
Hearing Impaired	72	454	97	53	47	55	476	105	42	58	16	546	99	31	69

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Learning Disability	10	587	75	10	90	6	623	32	.	100	3	601	121	33	67
Multiple-Disability	82	419	75	78	22	84	445	87	68	32	25	474	121	68	32
Other Health Impaired	59	496	94	29	71	43	527	73	21	79	15	562	74	13	87
Orthopedically Impaired	59	478	82	46	54	51	501	80	35	65	11	516	118	45	55
Speech or Language Impaired	887	494	64	31	69	557	501	69	34	66	135	506	61	35	65
Traumatic Brain Injury	3	571	70	.	100	5	494	36	40	60	4	496	75	50	50
Visually Impaired															
TOTAL	1230	487	75	35	65	1048	504	81	34	66	371	512	88	36	64

Exhibit 24: Performance by Grade-Band Form and Student Age—Mathematics

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	389	477	70	49	51	.	.	.	.	.	.	.	.	.	.
9	425	491	73	39	61	.	.	.	.	.	.	.	.	.	.
10	401	498	77	39	61	1	409	.	100	.	.	.	.	.	.
11	7	591	58	.	100	353	494	80	46	54	.	.	.	.	.
12	2	340	67	100	.	336	502	80	39	61	.	.	.	.	.
13	2	519	45	50	50	353	513	87	39	61	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	1	575	.	.	100
15	.	.	.	.	.	.	.	.	.	.	369	500	82	41	59
Blank	1	514	.	.	100	.	.	.	.	.	.	.	.	.	.
STUDENT'S ETHNICITY															
African American	644	494	75	40	60	534	514	86	38	62	204	503	88	41	59
White/Asian	5	510	45	20	80	.	.	.	.	.	.	.	.	.	.
Other	3	513	75	33	67	4	486	39	75	25	.	.	.	.	.
African American/American Indian	3	489	48	33	67	.	.	.	.	.	1	590	.	.	100
American Indian	5	491	154	40	60	3	499	59	33	67	.	.	.	.	.
Asian	8	500	46	25	75	12	481	44	50	50	2	481	27	50	50
Hawaiian/Pacific Islander	1	369	.	100	.	1	492	.	.	100	1	510	.	.	100
Hispanic	47	482	63	49	51	35	484	68	49	51	3	553	68	.	100
White	493	483	73	45	55	442	493	79	45	55	158	497	74	42	58
White/African American	18	495	94	33	67	10	505	101	30	70	1	464	.	100	.
White/American Indian	.	.	.	.	.	2	556	90	.	100	.	.	.	.	.
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
STUDENT'S GENDER															
Female	404	483	79	45	55	355	502	81	41	59	136	492	83	38	62
Male	823	492	72	40	60	688	504	83	41	59	234	506	80	43	57
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Pre-functional	31	471	61	52	48	23	486	53	52	48	2	508	11	.	100
Beginner	1	526	.	.	100	1	499	.	.	100	.	.	.	.	.
Intermediate	.	.	.	.	.	2	506	42	50	50	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	1	439	.	100	.	1	520	.	.	100	.	.	.	.	.
Title III Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English speaker I	.	.	.	.	.	1	472	.	100	.	.	.	.	.	.
English speaker II	1188	490	74	41	59	1012	504	83	41	59	363	501	82	41	59
Pre-functional Waiver	1	458	.	100	.	.	.	.	.	.	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	1	601	.	.	100	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	5	460	54	60	40	2	491	49	50	50	5	466	67	60	40
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	386	478	73	47	53	329	485	75	48	52	113	484	85	48	52
Free	742	496	75	38	62	627	513	83	38	62	224	510	80	36	64
Reduced	97	480	64	49	51	87	504	93	40	60	30	498	73	47	53
Blank	2	474	3	50	50	.	.	.	.	.	3	421	35	100	.
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)															
Severe Mental Disability	128	393	64	95	5	108	393	62	99	1	35	390	84	94	6
Moderate Mental Disability	371	475	45	53	47	439	490	59	46	54	181	497	62	41	59
Mild Mental Disability	325	546	57	9	91	245	573	66	7	93	66	564	64	11	89
Autism	252	483	56	42	58	152	489	66	47	53	45	489	75	49	51
Deaf/Blindness	3	368	163	67	33	3	491	59	67	33	4	520	47	25	75
Emotional Disability	7	565	93	14	86	4	626	28	.	100	.	.	.	.	.
Hearing Impaired	72	452	95	61	39	55	476	104	49	51	16	517	105	38	63
Learning Disability	10	602	77	10	90	6	672	38	.	100	3	529	73	33	67
Multiple-Disability	83	423	73	77	23	83	446	88	72	28	25	452	122	68	32

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Other Health Impaired	59	496	90	41	59	43	526	82	35	65	15	547	62	13	87
Orthopedically Impaired	59	471	73	51	49	51	498	75	45	55	10	494	76	60	40
Speech or Language Impaired	886	495	64	39	61	554	499	71	42	58	135	499	54	42	58
Traumatic Brain Injury	3	573	70	.	100	5	509	34	20	80	4	461	136	25	75
Visually Impaired	49	423	89	71	29	32	445	95	66	34	7	469	127	71	29
TOTAL	1227	489	74	42	58	1043	503	82	41	59	370	501	82	41	59

Exhibit 25: Performance by Grade-Band Form and Student Age—Science

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	224	474	68	41	59	.	.	.	.	.	.	.	.	.	.
9	419	490	68	31	69	.	.	.	.	.	.	.	.	.	.
10	229	495	68	29	71	1	413	.	100	.	.	.	.	.	.
11	5	566	32	.	100	205	497	84	44	56	.	.	.	.	.
12	2	347	113	100	.	333	504	75	37	63	.	.	.	.	.
13	1	445	.	100	.	220	505	86	34	66	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	1	531	.	.	100
15	.	.	.	.	.	.	.	.	.	.	367	504	90	49	51
STUDENT'S ETHNICITY															
African American	465	487	72	31	69	376	509	81	37	63	203	506	99	49	51
White/Asian	3	507	50	33	67	.	.	.	.	.	.	.	.	.	.
Other	3	507	55	33	67	4	495	22	50	50	.	.	.	.	.
African American/American Indian	2	486	58	50	50	.	.	.	.	.	1	642	.	.	100
American Indian	3	584	31	.	100	3	492	97	33	67	.	.	.	.	.
Asian	5	477	11	40	60	6	474	29	50	50	2	437	107	50	50
Hawaiian/Pacific Islander	1	370	.	100	.	1	504	.	.	100	1	495	.	100	.
Hispanic	34	476	53	38	62	27	477	82	37	63	3	576	143	33	67
White	352	487	67	35	65	330	497	80	40	60	157	501	77	50	50
White/African American	12	497	69	33	67	10	531	96	30	70	1	462	.	100	.
White/American Indian	.	.	.	.	.	2	531	113	50	50	.	.	.	.	.
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
STUDENT'S GENDER															
Female	279	482	72	35	65	250	496	81	39	61	135	494	88	52	48
Male	601	489	67	32	68	509	506	80	38	62	233	510	91	48	52
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															
Pre-functional	25	464	63	44	56	12	477	69	33	67	2	510	3	.	100

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Beginner	1	527	.	.	100	1	514	.	.	100	.	.	.	.	.
Intermediate	.	.	.	.	.	2	530	55	.	100	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	1	462	.	100	.	1	501	.	.	100	.	.	.	.	.
Title III Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English speaker I	.	.	.	.	.	1	488	.	100	.	.	.	.	.	.
English speaker II	847	488	69	33	67	739	503	81	38	62	361	505	91	49	51
Pre-functional Waiver	1	498	.	.	100	.	.	.	.	.	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	1	586	.	.	100	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	5	463	67	40	60	2	505	33	50	50	5	489	85	60	40
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	281	479	70	35	65	245	485	76	47	53	113	486	96	56	44
Free	529	492	68	31	69	456	513	81	33	67	222	515	87	45	55
Reduced	68	477	68	43	57	58	495	86	38	62	30	502	83	50	50
Blank	2	508	22	.	100	.	.	.	.	.	3	430	34	100	.
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)															
Severe Mental Disability	93	391	66	92	8	81	387	67	96	4	35	372	87	97	3
Moderate Mental Disability	264	480	48	34	66	307	496	65	39	61	180	506	69	49	51
Mild Mental Disability	225	535	41	4	96	181	566	51	5	95	65	569	67	15	85
Autism	189	481	54	38	62	118	482	65	56	44	45	484	89	67	33
Deaf/Blindness	2	404	148	50	50	2	508	38	50	50	4	548	62	25	75
Emotional Disability	5	553	21	.	100	3	566	27	.	100	.	.	.	.	.
Hearing Impaired	53	459	71	53	47	41	488	96	44	56	16	533	99	38	63
Learning Disability	7	555	64	14	86	6	628	55	.	100	3	554	82	33	67
Multiple-Disability	58	415	81	76	24	61	445	97	69	31	25	455	130	68	32
Other Health Impaired	41	490	94	32	68	31	534	75	19	81	15	548	52	27	73
Orthopedically	42	475	80	43	57	31	496	66	35	65	10	490	100	60	40

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Impaired															
Speech or Language Impaired	632	493	58	28	72	407	502	73	39	61	135	504	62	52	48
Traumatic Brain Injury	2	611	85	.	100	4	494	10	25	75	4	486	96	50	50
Visually Impaired	38	439	69	66	34	22	477	98	55	45	7	490	120	71	29
TOTAL	880	487	69	33	67	759	503	80	38	62	368	504	90	49	51

Exhibit 26: Performance by Grade-Band Form and Student Age—Social Studies

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	213	480	74	52	48	.	.	.	.	.	.	.	.	.	.
9	416	498	75	43	57	.	.	.	.	.	.	.	.	.	.
10	208	499	89	41	59	.	.	.	.	.	.	.	.	.	.
11	5	600	97	20	80	195	497	89	51	51	.	.	.	.	.
12	1	431	.	100	.	332	512	85	43	43	.	.	.	.	.
13	2	501	65	50	50	196	524	97	38	38	.	.	.	.	.
Blank	1	537	.	.	100	.	.	.	.	.	.	.	.	.	.
STUDENT'S ETHNICITY															
African American	460	496	82	45	55	373	522	90	38	38	.	.	.	.	.
White/Asian	2	482	11	100	.	.	.	.	.	.	.	.	.	.	.
Other	2	532	40	.	100	1	493	.	100	100	.	.	.	.	.
African American/American Indian	2	522	11	.	100	.	.	.	.	.	.	.	.	.	.
American Indian	4	482	188	50	50	1	506	.	.	.	.	.	.	.	.
Asian	8	484	44	63	38	9	491	28	56	56	.	.	.	.	.
Hawaiian/Pacific Islander	1	394	.	100	.	.	.	.	.	.	.	.	.	.	.
Hispanic	32	498	63	41	59	25	481	83	60	60	.	.	.	.	.
White	323	493	75	46	54	305	501	90	49	49	.	.	.	.	.
White/African American	12	488	88	33	67	7	491	100	57	57	.	.	.	.	.
White/American Indian	.	.	.	.	.	2	498	68	50	50	.	.	.	.	.
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
STUDENT'S GENDER															
Female	286	487	82	45	55	243	509	95	42	42	.	.	.	.	.
Male	560	498	77	45	55	480	512	87	45	45	.	.	.	.	.
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															
Pre-functional	21	472	67	62	38	20	499	62	55	55	.	.	.	.	.
Beginner	1	549	.	.	100	.	.	.	.	.	.	.	.	.	.
Intermediate	.	.	.	.	.	2	550	81	50	50	.	.	.	.	.

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	1	479	.	100	.	1	490	.	100	100	.	.	.	.	.
Title III Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English speaker I	.	.	.	.	.	1	470	.	100	100	.	.	.	.	.
English speaker II	821	495	79	44	56	697	511	91	43	43	.	.	.	.	.
Pre-functional Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	1	591	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	2	492	54	50	50	1	551	.	.	.	.	.	.	.	.
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	262	491	84	48	52	229	498	89	49	49	.	.	.	.	.
Free	511	498	79	42	58	434	519	90	40	40	.	.	.	.	.
Reduced	72	488	62	53	47	60	503	88	48	48	.	.	.	.	.
Blank	1	454	.	100	.	.	.	.	.	.	.	.	.	.	.
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)															
Severe Mental Disability	93	384	78	96	4	75	379	74	99	99	.	.	.	.	.
Moderate Mental Disability	246	484	53	56	44	306	506	62	47	47	.	.	.	.	.
Mild Mental Disability	234	553	49	10	90	162	590	69	9	9	.	.	.	.	.
Autism	170	489	61	53	47	111	492	73	57	57	.	.	.	.	.
Deaf/Blindness	1	577	.	.	100	2	515	21	50	50	.	.	.	.	.
Emotional Disability	3	581	46	.	100	4	626	52	.	.	.	.	.	.	.
Hearing Impaired	47	458	89	60	40	42	466	115	55	55	.	.	.	.	.
Learning Disability	6	574	43	.	100	3	642	44	.	.	.	.	.	.	.
Multiple-Disability	55	425	88	80	20	57	443	90	72	72	.	.	.	.	.
Other Health Impaired	40	502	75	38	63	28	524	69	36	36	.	.	.	.	.
Orthopedically Impaired	40	494	78	43	58	36	512	90	36	36	.	.	.	.	.
Speech or Language Impaired	612	503	65	42	58	389	507	78	46	46	.	.	.	.	.

	Elementary					Middle					High				
	N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)		N	Scale Score		Ach. Level (%)	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Traumatic Brain Injury	2	548	32	.	100	5	507	18	60	60	.	.	.	.	.
Visually Impaired	33	437	91	67	33	22	425	106	73	73	.	.	.	.	.
TOTAL	846	495	79	45	55	723	511	90	44	44	.	.	.	.	.

Chapter 8: Validity

Content Validity

One source of evidence for the content validity of the South Carolina Alternate Assessment was obtained through independent alignment studies. The University of North Carolina at Charlotte (UNCC) conducted studies of the alignment of (a) ASMGs to grade-level curriculum standards and (b) SC-Alt items to the ASMGs that they targeted. This was a pilot study conducted by Flowers, Browder, Wakeman, and Karvonen with UNCC through the National Alternate Assessment Center (NAAC). (South Carolina is a member state of the NAAC.) A second independent study of ELA and mathematics was completed by the South Carolina Education Oversight Committee (EOC; 2008a) as required by the state Education Accountability Act of 1998 (EAA). The EOC approved the ELA and mathematics content areas on February 28, 2008. The UNCC alignment study results for the English language arts and mathematics assessments are reported in detail in Flowers, Browder, Wakeman, and Karvonen (2006a). The results of the alignment studies for the ELA and mathematics assessments indicate that

> The state has evidence supporting alignment for its measurement guidelines and alternate assessment based on all seven criteria. We conclude that overall this is an alternate assessment system that links to the grade level content. Some areas for consideration in further development of the system are noted related to balance of content. (p. 7)

The alignment study results for the science assessment are reported in detail in Flowers, Browder, Wakeman, and Karvonen (2006b) and in an addendum dated December 21, 2007. The results of the alignment study for the science assessment indicate that

> The strength of the South Carolina science Alternate Assessment was that nearly all of the content was academic science content (98%). This is especially notable given that the alternate assessment tasks included items accessible to students at all symbolic levels. In contrast, the degree of alignment of AA tasks/items to grade-level standards was lower than those found in the alignment of ELA and mathematics. This difference could be due to the fact that the state's science grade-level standards changed during the development of the science AA. Another challenge was that the state had linked its alternate assessment tasks to the state standards and not directly to the measurement guidelines, creating a tough challenge to demonstrating alignment…. Our work with other states suggests that science may typically be the area rated as having the weakest alignment. (p. 4).

SCDE reviewed the initial science alignment study and determined that one source of some misalignment had resulted from the linking of some items to multiple standards and indicators in the alignment document provided by SCDE. During the Science Content Review Committee meeting, some members recommended adding additional indicators to align to some items. The intent of these recommendations focused more on instruction and demonstrating that instruction could include multiple standards and indicators. However, the alignment study team considered only the first two standards aligned to each item. In some cases, the first two

standards were not necessarily the most appropriate. SCDE prioritized the standards and indicators and resubmitted the documentation for an additional study. From this review, completed December 21, 2007 (Flowers, Browder, Wakeman, and Karvonen, 2007), 163 of 173 items were rated as academic. Of the 10 items listed as nonacademic, six were rated as foundational (p.1). SCDE is currently addressing the items that were rated as having no content centrality by developing replacement items for new forms.

At the time of the alignment study for ELA and mathematics by Dr. Flowers and colleagues, the design of the SC-Alt was envisioned as a single assessment across grade levels. This design changed to a grade-band assessment following the study; however, the information provided from the alignment study was used to identify items with alignment difficulty, and these items were omitted from the operational grade-band test forms. Information from the review along with teacher comments was also used during item data review as part of the decision-making process regarding inclusion of items in the assessment.

A second independent review of the alignment of the science assessment was conducted by the Education Oversight Committee (EOC; 2008b) The EOC approved the elementary and middle school science alternate assessment on August 12, 2008. The EOC alignment findings were based on the review of two sets of studies of the SC-Alt:

- studies of the alignment between the SC-Alt Science assessment and the state academic standards conducted by University of North Carolina-Charlotte and Western Carolina University professors of curriculum and special education, in cooperation with the South Carolina State Department of Education (SCDE) and the National Alternate Assessment Center (Flowers, Browder, Wakeman, & Karvonen, December 2006; January 2007; December 2007); and
- a technical review of the task and item data from the 2007 test administration conducted by a professor of educational research and assessment at the University of South Carolina.

Copies of the reports of the EOC reviews and findings are available in their entirety from the SCDE. Based on this review, the EOC identified a number of strengths of the SC-Alt science assessment that were noted in the final report:

- The assessment provides accountability and information for instructional improvement for students with significant cognitive disabilities who would not otherwise be assessed in the state testing programs, even with test accommodations and modifications.
- The assessment is intended to be aligned with the same grade level academic standards as for all students, although at levels of complexity appropriate for the diversity of cognitive functioning observed among students with significant cognitive disabilities.
- The assessment format allows students to respond to the items using the communication modes the student uses during instruction, such as oral response, pointing, use of eye gaze, use of a response card, sign language, or an augmentative communication device.
- The procedures for placing the student at the appropriate level for beginning each assessment reduces student fatigue and maximizes students' opportunities to show their highest performance;
- The items in the assessment have a wide range of difficulty and the test is moderately able to discriminate between high and low levels of performance.

The EOC report noted that while 96% of the items were found to be aligned to science inquiry standard indicators, the alignment of the items to content standards was 78%, falling short of an expectation for successful alignment of 90% set by the original evaluators. The EOC recommended that the SCDE review the alignment of the SC-Alt science items to the grade-level standards and identify items needing revision or replacement. During 2008, SCDE and AIR developed five new tasks consisting of 32 items to be used to replace poorly aligned items. An independent review of the alignment of the new items by the Center for Research on Education (2009) found that 98% of the new items were aligned to grade-level content standard indicators.

Convergent and Discriminant Validity

According to Critical Element 4.1(e) of the federal peer review and Standard 1.14 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), it is desirable, if not necessary, to provide evidence of convergent and discriminant validity. One common method for examining this aspect of validity is with a multitrait-multimethod matrix (MTMM) (Campbell & Fiske, 1959).

Campbell and Fiske (1959) proposed the MTMM matrix design as a tool for the study of convergent and discriminant validity in psychological measurement. The MTMM matrix employs a crossed factorial measurement design of traits and methods to reveal these types of validity in comparison:

- Large correlations on validity diagonals (i.e., same trait and different methods) indicate convergent validity.
- Low correlations in the heterotrait-monomethod blocks indicate discriminant validity and the absence of method effects.
- Low correlations in the heterotrait-heteromethod blocks also indicate discriminant validity.

Selection of Traits and Methods

The student's abilities in each of the subjects—ELA, mathematics, science, and social studies—make up the four traits for the MTMM study. Two methods are considered for assessing these traits: the Student Placement Questionnaire (SPQ) as a structured teacher rating of student ability and the SC-Alt scale score as an IRT score of the student's responses to the set of presented test items. In other words, the two methods contrast test scores of student performance with expert (or teacher) ratings. With four traits and two methods, the MTMM correlation matrix is of order 8. Note that the grade 10 assessment did not include a social studies component; therefore, the MTMM matrix for grade 10 has only six rows and columns.

Results

MTMM matrices were computed separately for each grade band. The results are given in Exhibit 27 through Exhibit 29. Pearson correlations are used, with pairwise deletion of missing data. For each matrix, the minimum pairwise sample size is indicated. *P*-values of individual correlation coefficients are not reported since *all* correlations are significant ($p < 0.05$).

Exhibit 27: MTMM Matrix, Scale Scores with SPQ Scores, Grades 3–5

	IRT Scale Scores				**SPQ Scores**			
	ELA	Math	Science	Social Studies	ELA	Math	Science	Social Studies
ELA_Scale	(0.91)							
Math_Scale	*0.90*	(0.90)						
Science_Scale	*0.89*	*0.88*	(0.90)					
Social_Scale	*0.89*	*0.88*	*0.90*	(0.89)				
ELA_SPQ	**0.78**	0.78	0.77	0.77	(0.95)			
Math_SPQ	0.76	**0.79**	0.74	0.76	*0.91*	(0.97)		
Science_SPQ	0.72	0.74	**0.72**	0.69	*0.86*	*0.89*	(0.96)	
Social_SPQ	0.77	0.78	0.77	**0.77**	*0.89*	*0.88*	*0.89*	(0.96)
Minimum pairwise N: 880								

Exhibit 28: MTMM Matrix, Scale Scores and SPQ Scores, Grades 6–8

	IRT Scale Scores				**SPQ Scores**			
	ELA	Math	Science	Social Studies	ELA	Math	Science	Social Studies
ELA_Scale	(0.91)							
Math_Scale	*0.91*	(0.91)						
Science_Scale	*0.89*	*0.89*	(0.91)					
Social_Scale	*0.89*	*0.89*	*0.91*	(0.87)				
ELA_SPQ	**0.80**	0.78	0.79	0.77	(0.96)			
Math_SPQ	0.81	**0.82**	0.79	0.78	*0.91*	(0.97)		
Science_SPQ	0.77	0.77	**0.77**	0.73	*0.88*	*0.92*	(0.97)	
Social_SPQ	0.79	0.78	0.79	**0.78**	*0.90*	*0.90*	*0.92*	(0.97)
Minimum pairwise N: 723								

Exhibit 29: MTMM Matrix, Scales Scores and SPQ Scores, Grade 10

	IRT Scale Scores			**SPQ Scores**		
	ELA	Math	Science	ELA	Math	Science
ELA_Scale	(0.91)					
Math_Scale	*0.91*	(0.91)				
Science_Scale	*0.91*	*0.93*	(0.88)			
ELA_SPQ	**0.76**	0.75	0.74	(0.96)		
Math_SPQ	0.75	**0.75**	0.72	*0.93*	(0.98)	
Science_SPQ	0.75	0.73	**0.73**	*0.89*	*0.90*	(0.98)
Minimum pairwise N: 337						

In each MTMM table, the convergent validity coefficients (correlations between measurements of the same trait using different methods) are marked in bold. These convergent validity coefficients range from 0.72 to 0.82 and certainly fall into an acceptable range. These high correlations indicate good validity for the SPQ. The above three exhibits indicate that the SPQ and the actual test are essentially measuring the same trait and that the SPQ is a good indicator of performance on the test.

The entries in the monomethod triangles (correlations between measurements of different traits using the same method) are set in italics. These correlations coefficients range between 0.88 and 0.93 for IRT scale scores and between 0.86 and 0.93 for SPQ scores. These high correlations indicate the presence of method variance. However, this is to be expected because the SPQ was not developed to measure the trait; instead, it only indicates the starting task on the test for measuring the trait. Such a result of high correlations in the monomethod triangles is not uncommon in MTMM studies (Fiske, 1995). Specific conditions offer themselves as causes for the present scenario. First, the different scale types—number-correct rating scales versus IRT

scales of behavioral tests—are in themselves a source of method variation. Second, the SPQ's "can do" questions draw on the teacher's memory of a student's possible performance over the long term and are apt to differ in quality and veracity. Third, the IRT scale scores for the three subjects reflect the student's performance in the testing situation and are subject to the student's form on the testing day.

The heterotrait-heteromethod correlations appear in the tables in regular type. These correlation coefficients fall into the same range as the convergent validity coefficients, with values from 0.72 to 0.79. To confirm discriminant validation, the heterotrait-heteromethod correlations should be smaller than the convergent validity coefficients. Instead, these MTMM matrices support the notion that that the three traits vary essentially on just a single dimension. Because the population of alternate assessment students is so *very* heterogeneous, the students' general level of cognitive functioning dominates the relationship among their scale scores.

The estimates of marginal reliability of the IRT scale scores are in parentheses on the main diagonal. These coefficients were discussed in the Section about Test Score Reliability (page 57 and following).

Validity of the Student Placement Questionnaire (SPQ)

AIR analyzed information from the 2008 administration regarding the agreement between SPQ recommended start points and the final observed start points as determined by a review of the 2008 item data. The purpose of the study was to determine the effectiveness of the SPQ in identifying the most appropriate starting task.

Administration of the SC-Alt uses the SPQ as a pre-assessment instrument to determine the most appropriate starting point in the assessment. The SPQ requires the teacher to evaluate the student on 12 to 15 "can do" statements addressing the student's skills and knowledge in each content area on the basis of the teacher's prior instructional knowledge of the student. A total score computed from the teacher's SPQ responses indicates the initial starting task for the assessment.

The instructions for using the SPQ require teachers to adjust the starting point below the SPQ recommended start point when the student is not successful on the first administered task. Alternately, after reviewing the assessment, some teachers may have judged that a student needed to start at a higher level than recommended by the SPQ. This result occurred almost exclusively when the SPQ recommended starting point was Task 1.

A summary of the results of the agreement between the SPQ recommended start points and the observed start points for each content area and grade-band form is presented in Exhibit 30. These results indicate that the agreement between the SPQ recommended start point and the observed start point was 92% for ELA, 92% for mathematics, and 88% for science administrations. Use of the SPQ pre-assessment score is only the first step in the procedure used by the test administrator in determining where the student should start the assessment. Since the test administrator is required to make adjustments based on the student's success on the first task, and these adjustments are reflected in the agreement rates, the SPQ appears to be working very effectively for targeting the first task to begin the assessment process.

The 2008 administration was the first operational administration for social studies. The agreement between the SPQ recommended start point and the observed start point for social studies was 77%. This lower agreement level for social studies was apparently due to misstated SPQ instructions on the spring 2008 student score form.

Exhibit 30: Agreement Between SPQ and Observed Start Points

Subject	Grade-Band	SPQ Recommended Starting Task	Starting Task Consistent with SPQ	Lower Start Task than Recommended	Higher Start Task than Recommended	Non-standard Start Task	Incomplete SPQ	Percentage Inconsistent with SPQ
ELA	Elementary	1	370	0	5	2	0	1.86%
		3	269	10	2	7	0	6.60%
		6	496	13	0	2	0	2.94%
		Total	1135	23	7	11	49	7.35%
	Middle	1	267	0	2	0	0	0.74%
		3	174	13	1	2	0	8.42%
		6	531	8	0	2	0	1.85%
		Total	972	21	3	4	39	6.45%
	High	1	86	0	0	0	0	0.00%
		3	57	12	1	0	0	18.57%
		6	180	6	0	1	0	3.74%
		Total	323	18	1	1	29	13.17%
	ELA Total		**2430**	**62**	**11**	**16**	**117**	**7.81%**
Math	Elementary	1	346	0	4	0	0	1.14%
		3	318	14	2	3	0	5.64%
		6	465	9	0	4	0	2.72%
		Total	1129	23	6	7	58	7.69%
	Middle	1	257	0	1	0	0	0.39%
		3	200	17	3	1	0	9.50%
		6	499	9	0	2	0	2.16%
		Total	956	26	4	3	46	7.63%
	High	1	90	0	0	0	0	0.00%
		3	72	7	0	1	0	10.00%
		6	165	10	0	3	0	7.30%
		Total	327	17	0	4	23	11.86%

Subject	Grade-Band	SPQ Recommended Starting Task	Starting Task Consistent with SPQ	Lower Start Task than Recommended	Higher Start Task than Recommended	Non-standard Start Task	Incomplete SPQ	Percentage Inconsistent with SPQ
Math Total			**2412**	**66**	**10**	**14**	**127**	**8.25%**
Science		1	312	0	7	11	0	5.45%
		3	198	6	1	0	0	3.41%
		6	241	48	0	0	0	16.61%
	Elementary	Total	751	54	8	11	54	14.46%
		1	232	0	1	0	0	0.43%
		3	133	7	1	2	0	6.99%
		6	331	2	0	0	0	0.60%
	Middle	Total	696	9	2	2	46	7.81%
		1	111	0	1	0	0	0.89%
		3	50	3	1	0	0	7.41%
		6	162	3	0	3	0	3.57%
	High	Total	323	6	2	3	35	12.47%
Science Total			**1770**	**69**	**12**	**16**	**135**	**11.59%**
Social Studies		1	229	0	2	0	0	0.87%
		3	121	10	2	58	0	36.65%
		6	292	7	0	75	0	21.93%
	Elementary	Total	642	17	4	133	49	24.02%
		1	163	0	2	1	0	1.81%
		3	76	9	4	34	0	38.21%
		6	317	9	0	59	0	17.66%
	Middle	Total	556	18	6	94	41	22.24%
Social Studies Total			**1198**	**35**	**10**	**227**	**90**	**23.21%**
Grand Total			**7810**	**232**	**43**	**273**	**469**	**11.52%**

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC.

American Institutes for Research. (2007). *South Carolina Alternate Assessment Spring 2007 Standard Setting Technical report.* Washington, DC: Author**.**

American Institutes for Research (2008) *South Carolina Alternate Assessment (SC-Alt): Technical report for English Language Arts and Mathematics Field Test Administration, Spring 2006*. Washington, DC: Author

American Institutes for Research and South Carolina Department of Education (2008) *South Carolina's Alternate Assessment, SC-Alt Spring 2007 Operational Administration*. Washington, DC: American Institutes for Research

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Center for Research on Education (2009). *Alternate assessment alignment study of replacement items report to the South Carolina State Department of Education*, Columbia, SC: South Carolina State Department of Education.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage Publications.

Dorans, N., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23,* 355–368.

Education Oversight Committee (2008a). *Review of the SC-Alternate English language arts and mathematics assessments report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education

Education Oversight Committee (2008b). *Review of the SC-Alternate science assessment report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education

Ferrara, S., Perie, M., & Johnson, E. (2008). Matching the judgmental task with standard setting panelist expertise: The Item-Descriptor (ID) Matching procedure. *Journal of Applied Testing Technology*, 9, 1-22.

Fiske, D. W. (1995) Reprise, new themes, and steps forward. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2006a). *Alternate assessment alignment pilot study report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2006b). *Science alternate assessment alignment pilot study report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2007). *Science alternate assessment alignment pilot study report to the South Carolina State Department of Education, Addendum, December 21, 2007*. Columbia, SC: South Carolina State Department of Education.

Huynh, H. (2003, August). *Technical memorandum for computing standard error in Bookmark standard setting* [The South Carolina PACT 2003 Standard Setting Support Project]. Columbia, SC: University of South Carolina.

Kleinert, H. L., and Kearns, J. F. (2001). *Alternate assessment: Measuring outcomes and supports for students with disabilities.* Baltimore, MD: Brookes Publishers.

Linacre, J. M., & Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370. Available at http://www.rasch.org/rmt/rmt83.htm.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.

National Alternate Assessment Center (2005) *Designing from the Ground Floor*. Conference Pre-Session, Chief Council for State School Officers' Large-Scale Assessment Conference. San Antonio, TX: June 17-18, 2005.

Rudner, L.M. & Schafer, W.D. (2001) Reliability (ERIC Digest). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation (ERIC Document Reproduction Service No. ED458213). Retrieved May 17, 2009 from http://www.ericdigests.org/2002-2/reliability.htm

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, *28*(3), 234–247.

Spitzer, R., Cohen, J., Fleiss, J., & Endicott, J. (1967). Quantification of agreement in psychiatry diagnosis: A new approach. *Archives of General Psychiatry*, *17*, 83–87.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, *30*, 233–251.

Zwick, R., & Thayer, D. T. (1996). Evaluating the magnitude of differential item functioning in polytomous items. *Journal of Educational and Behavioral Statistics*, *21*(3), 187–201.

Appendices

Appendix A: Assignment of Tasks to Grade-Band Forms for the Spring 2008 Administration

All tasks in each SC-Alt grade-band assessment align with the Assessment Standards and Measurement Guidelines (ASMGs) in that grade band. Because adjacent grade-band score scales are linked psychometrically, some tasks in each grade-band assessment align with ASMGs in both adjacent grade bands. In turn, these separate grade-band ASMGs link to separate grade-level performance standards for the appropriate grades.

All items in linking tasks are developed to be appropriate for students in both adjacent grade bands. In some cases (e.g., some tasks in ELA), the ASMGs to which linking tasks align are equivalent for two adjacent grade bands. However, the grade-level performance standards to which the ASMGs are linked *do* differ across the adjacent grade bands. In all content areas and for all grade bands, Descriptions of Achievement Levels (DALs) are specific to each grade band and differ across grade bands.

2008 Operational Field-Test Designs to Support Psychometric Linking of Grade-Band Score Scales

To provide data to link all grade-band assessments onto a vertical scale, linking tasks were repeated in adjacent grade assessments. For example, five of the tasks that appeared in the ELA grades 3–5 assessment also appeared in the ELA grades 6–8 assessment. Those five linking tasks and the seven unique tasks made up the 12 tasks in the ELA grades 3–5 assessment. The ELA grades 6–8 assessment included the five linking tasks from the 3–5 assessment, five linking tasks that also appeared in the grades 9–12 assessment, and two unique tasks. This "linking upward" design ensures that students were assessed on ASMGs aligned with their current grade placement or previous grades.

English Language Arts Assessment

The ELA assessment covered ASMGs in reading, writing, and communication. The design included 12 tasks for each of three grade-band assessments and required the development of a total of 26 tasks. The design for the ELA assessment for spring 2008 appears in Exhibit 31.

Exhibit 31: Numbers of Tasks in Each Operational Grade-Band Assessment, ELA

<table>
<tr><th>Grade Band</th><th>Unique Tasks</th><th colspan="2">Linking Tasks</th><th>Total for Operational Test</th></tr>
<tr><td></td><td colspan="4">Tasks in Each Grade-Band Assessment Test Booklet</td></tr>
<tr><td>10</td><td>7</td><td rowspan="2">5</td><td>—</td><td>12</td></tr>
<tr><td>6–8</td><td>2</td><td rowspan="2">5</td><td>12</td></tr>
<tr><td>3–5</td><td>7</td><td>—</td><td>12</td></tr>
<tr><td></td><td></td><td colspan="2"></td><td></td></tr>
<tr><td></td><td colspan="4">Tasks to Be Included</td></tr>
<tr><td>All grades</td><td>16</td><td colspan="2">10</td><td>26</td></tr>
</table>

Mathematics Assessment

The mathematics assessment covered the mathematics ASMGs. The design included 12 tasks for each of three grade-band assessments and required the development of a total of 22 tasks. The design for the mathematics assessment for spring 2008 appears in Exhibit 32.

Exhibit 32: Numbers of Tasks in Each Operational Grade Band Assessment, Mathematics

<table>
<tr><th>Grade Band</th><th>Unique Tasks</th><th colspan="3">Linking Tasks</th><th>Total for Operational Test</th></tr>
<tr><td></td><td colspan="5">Tasks in Each Grade-Band Assessment Test Booklet</td></tr>
<tr><td>10</td><td>6</td><td rowspan="2">2</td><td>—</td><td rowspan="3">4</td><td>12</td></tr>
<tr><td>6–8</td><td>2</td><td rowspan="2">4</td><td>12</td></tr>
<tr><td>3–5</td><td>4</td><td>—</td><td>12</td></tr>
<tr><td></td><td></td><td colspan="3"></td><td></td></tr>
<tr><td></td><td colspan="5">Tasks to Be Included</td></tr>
<tr><td>All grades</td><td>12</td><td colspan="3">10</td><td>22</td></tr>
</table>

Science Assessment

The science assessment covered the science ASMGs. The design included 12 tasks for each of three grade-band assessments and required the development of a total of 27 tasks. The design for the science assessment for spring 2008 appears in Exhibit 33.

Exhibit 33: Numbers of Tasks in Each Operational Grade-Band Assessment, Science

<table>
<tr><th>Grade Band</th><th>Unique Tasks</th><th colspan="2">Linking Tasks</th><th>Total for Operational Test</th></tr>
<tr><td></td><td colspan="4">Tasks in Each Grade-Band Assessment Test Booklet</td></tr>
<tr><td>10</td><td>8</td><td rowspan="2">4</td><td>—</td><td>12</td></tr>
<tr><td>6–8</td><td>3</td><td rowspan="2">5</td><td>12</td></tr>
<tr><td>3–5</td><td>7</td><td>—</td><td>12</td></tr>
<tr><td></td><td></td><td colspan="2"></td><td></td></tr>
<tr><td></td><td colspan="4">Tasks to Be Included</td></tr>
<tr><td>All grades</td><td>18</td><td colspan="2">9</td><td>27</td></tr>
</table>

Social Studies Assessment

The social studies assessment covered the social studies ASMGs. The design included 12 operational and one field-test tasks for each of two grade-band assessments and required the development of a total of 20 tasks. The design for the science assessment for spring 2008 appears in Exhibit 34.

Exhibit 34: Numbers of Tasks in Each Operational/Field-Test Grade-Band Assessment, Social Studies

Grade Band	Unique Tasks	Linking Tasks	Total for Operational Test
Tasks in Each Grade-Band Assessment Test Booklet			
6–8	7	5 (+ 1 FT task)	13
3–5	7		13
Tasks to Be Included			
All grades	14	5 (+ 1 FT task)	20

Appendix B: Starting and Stopping Rules for Using the Student Placement Questionnaire

Directions for Determining the Starting and Concluding Tasks and Use of the Student Placement Questionnaire, Spring 2008

These directions guide you through the following steps:

- completing the Student Placement Questionnaire (SPQ),
- identifying the starting task in each content area,
- adjusting the starting task, if that becomes necessary,
- determining when to conclude the administration

Completing the Student Placement Questionnaire

The SPQ is designed to identify the most appropriate starting task for each of your students in each content area of SC-Alt. You will use the SPQ to identify the most appropriate starting task for each student in the SC-Alt assessments in English language arts, mathematics, science, and social studies. Answer each SPQ item as accurately as you can based on your experience in the classroom with this student.

The SPQs are located in the Student Answer Folder along with the areas for recording the student's scores on each SC-Alt task. **An example of a completed English Language Arts SPQ is included at the end of these instructions.**

Identifying the Starting Task for a Student in Each Content Area

1. Bubble in your responses to the SPQ questions.
2. After you respond to all items in the SPQ, identify the most appropriate starting task for this student following the steps on the SPQ. These are the steps:
    a. Count the number of bubbles you marked in each of the first three columns, and write the totals in the blocks under each column.
    b. In section 3 at the bottom of the page:
        i. Write the column totals in the appropriate blocks.
        ii. Multiply each total by the specified multiplier, and write the resulting totals in the blocks to the right.
        iii. Sum the three totals to obtain the total SPQ score. Write the SPQ score into the blocks and bubble in the SPQ score.

        Please check your work and complete the bubble grids for the total SPQ score.
    c. Find the total SPQ score in section 4 to determine the starting task for this student.

Administering the Starting Task and Completing the Administration

After you identify the starting task for this student using the SPQ, follow these directions to administer the starting task and complete the administration.

The SPQ provides the initial starting point for a student's administration. Each student must be administered a minimum of five tasks (including the starting task) if the student is started at Task 1 or a minimum of seven tasks if the student is started at Task 3 or Task 6. The minimum number of tasks and specific tasks that must be administered to each student for each starting level are specified in the table below.

Exhibit 35: Minimum Task Ranges to Be Administered

ELA, Mathematics, and Science	
Starting task	**Administer all items in at least these tasks**
Task 1	1–5
Task 3	3–9
Task 6	6–12

Social Studies	
Starting task	**Administer all items in at least these tasks**
Task 1	1–5, 13
Task 3	3–9, 13
Task 6	6–12, 13

Please note that the social studies
assessment has an additional 13th task that must be presented to each student.

It may be necessary to adjust the starting task based on the student's level of success on the first task. Also, the administration should be continued beyond the minimum number of tasks when the student is responding successfully.

When the Student Does Not Respond Successfully on the First Task

"Responding successfully" means getting at least three total points on a task. Each task has at least four items. Responding successfully would mean that a student received at least three total points for all the items combined. For example, a student may respond successfully by receiving three points on one item, two points on one item and one point on another item, or one point each on three different items. When a student does not receive three or more total points on a task, the student has not responded successfully on the task.

When a student is started at Task 3 or at Task 6 and does not respond successfully on the first task, the starting task was too difficult, and the teacher must restart the student at the next lower starting point. For example:

- If the student starts at Task 3 but cannot respond successfully on Task 3, restart the student at Task 1.
- If the student starts at Task 6 but cannot respond successfully on Task 6, restart the student at Task 3.

When a student is started at Task 1, no downward adjustment is possible, and the administration must progress through at least five tasks.

When to Conclude the Administration

If the student responds successfully on the last required task as specified in the table above, continue with the administration by administering the next task and subsequent tasks until the student no longer responds successfully on a task. By continuing the administration of subsequent tasks when the student is "responding successfully," you will provide the maximum opportunity for the student to demonstrate his or her knowledge and skills.

If the student does not respond successfully on the last required task or if at any point the student does not respond successfully on additional tasks (i.e., fails to obtain three or more points on the task), you may conclude the administration.

If you conclude the administration after administering the required tasks and when the student is no longer successful, you will not prolong the student's test administration unnecessarily and you will avoid any possible negative effects on the student.

Examples:

- Student A was started at Task 1 and administered Tasks 1–5. The student responded successfully on Task 5 and therefore was administered Task 6. The student responded successfully on Task 6 and was administered Task 7. The student did not respond successfully on Task 7, and the administration was concluded after Task 7.
- Student B was started at Task 3 and was administered Tasks 3–9. The student did not respond successfully on Task 9, and the administration was concluded after Task 9.

Additional Special Instructions for Social Studies Administrations:

There is one additional task to be administered to all students who take the social studies assessment. When the student has completed the concluding task, the test administrator should administer Task 13. This applies to both elementary and middle school forms.

SC - ALT STUDENT PLACEMENT QUESTIONNAIRE
ENGLISH LANGUAGE ARTS

(completed SPQ example)

Follow steps 1-4 to complete the SPQ and identify the starting task.

(1) Please darken the bubble (●) that corresponds to the most appropriate response for this student. Mark only one response for each item. Please mark a response for all items below. Use a No. 2 pencil only.

	Independently	With verbal/gestural prompting	With physical prompting/hand-over-hand	No, she/he cannot do this
In reading, can this student:				
1. Attend to text read aloud?	○	●	○	○
2. Recall details of text read aloud?	○	●	○	○
3. Recognize some high-frequency written words?	○	○	●	○
4. Draw conclusions or make inferences about texts?	○	○	○	●
In writing, can this student:				
5. Write his or her name using a pencil, name stamp, letter titles, or other means?	○	●	○	○
6. Use objects, pictures, and/or picture symbols to write in any format?	○	○	●	○
7. Copy, trace, or print letters?	○	●	○	○
8. Use oral language and/or letters and words to write?	○	○	○	●
In communicating, can this student:				
9. Listen (i.e., demonstrate receptive behavior) and respond?	●	○	○	○
10. Participate in conversations by responding appropriately?	○	●	○	○
11. Use language to express a preference, opinion, or viewpoint?	○	●	○	○
12. Recognize and understand the meaning of environmental signs (e.g., street signs, store signs, school signs)?	○	●	○	○
(2) Write in the total number of bubbles you marked in each column	1	7	2	
	Col. 1	*Col. 2*	*Col. 3*	

(3) Calculate the SPQ total score

(a) write the column totals from (2) in (a) below
(b) multiply and write the results in (b) below
(c) sum the results from (b) and write the sum in (c)

	(a)		(b)
Column 1 Total	1	x 3 =	3
		+	
Column 2 Total	7	x 2 =	14
		+	
Column 3 Total	2	x 1 =	2

(c) **Total SPQ Score** = 19

(4) Identify the starting task for this student using the SPQ total score from step (3).

If the total score is in this range	Start at this task	Administer all items in at least these tasks
0-11	Task 1	1-5
12-22	Task 3	3-9
23-36	Task 6	6-12

DO NOT WRITE IN THIS AREA

Appendix C: Scoring Audits and Analysis of Video-Rater Data from the Spring 2008 Operational Administration

A videotaping study was conducted to audit scoring accuracy for the spring 2008 administrations of the SC-Alt in ELA, mathematics, science, and social studies. **Scoring accuracy** refers to the degree to which teachers follow scaffolding and scoring directions correctly and assign correct scores to student responses. This appendix describes the sampling procedures, the identified sample of students, and the results for the attained sample of completed videotaped administrations.

Sampling Procedures

The sampling procedure was designed to include administrations from every school district and to be broadly representative of the range of student and test administrations. A sample of students was identified for videotaping so that (a) all districts implementing the SC-Alt would be required to videotape at least one student administration (all content areas) and (b) the total number of taped administrations per district would be based on the number of teachers involved in the assessment for each district. The sampling was by teacher and student within districts. One-third of the teachers within each district was randomly sampled to videotape the administration of one student. The number of teachers (and students) to be selected from each district is shown in Exhibit 36.

Exhibit 36: District-Based Sampling Targets for Video Rater Study

Total Number of Teachers per District	Number Required to Videotape
1–5	1
6–8	2
9–11	3
12–14	4
15–17	5
18–20	6
21–23	7
24–26	8
27–29	9
30–32	10
33–35	11
36–38	12
39–41	13
42–44	14
45 - 47	15
48 - 50	16
51 - 53	17
54 - 56	18
57 - 59	19

Based on this sampling plan and the numbers of pre-identified students coded for each district for the 2008 administration, the frequency distribution of students and teachers sampled per district was as follows:

- 1 teacher/student – 45 districts
- 2-5 teacher/students – 24 districts
- 6-10 teacher/students – 7 districts
- 11-15 teacher/students – 3 districts
- 16-19 teacher/students – 2 districts

The sampling of students and teachers was conducted from the January 2008 precode file, which was the pre-identification file for the spring 2008 SC-Alt administration. The sampling was conducted by SCDE, and the students identified for videotaping were flagged on the precode file sent to Measurement Incorporated (MI) for the production of materials and district notification. The numbers of students by form and disability sampled for videotaping are reported in Exhibit 37.

Exhibit 37: Stratified Sample of Students Identified for Videotaped Administrations

	Elementary		Middle		High School		Total	
	N	%	N	%	N	%	N	%
PRIMARY DISABILITY								
Severe Mental Disability	8	8.2	9	9.4	2	3.9	19	7.8
Moderate Mental Disability	29	29.6	32	33.3	21	41.2	82	33.5
Mild Mental Disability	16	16.3	24	25.0	8	15.7	48	19.6
Autism	25	25.5	13	13.5	10	19.6	48	19.6
Other Disabilities	20	20.4	18	18.8	10	19.6	48	19.6
TOTAL	98		96		51		245	

Videotaping Procedures

The district test coordinators for alternate assessment were provided rosters of the students identified for videotaping. The district materials included a packet of information for each teacher that included the following:

- a videotaping student roster identifying the student,
- information on the purpose of the videotaping and instructions for how to conduct the videotaped administrations,
- a videotaping student information form,
- bar code labels for positive identification and linking of the videotapes and the student information to the SC-Alt assessment data file, and
- directions for the packaging and return of materials.

The communications to both the teachers and the district test coordinators emphasized the importance of completing the videotaped administrations, provided contact information for questions or concerns, and asked for notification of SCDE if there were a problem in completing a videotaped administration for a particular student. Districts notified SCDE about a small number of students who either could not be assessed (e.g., because the students had moved, the parents did not consent to videotaping, or the students were not going to be assessed with the SC-Alt) or for whom the videotaping was inappropriate or extremely difficult to implement (e.g., medical homebound students or incompatible student behavior due to taping).

Most students who were deleted from the video sample list by SCDE were replaced by another student with the same teacher or in a few cases by identifying a different teacher and student. The replacement students were selected to match the grade-span form and disability of the original students as closely as possible. As a result of notifications by districts, the SCDE deleted 30 students from the original sample and instructed districts to videotape 17 replacement students.

Analysis of Video Rater Data

The total number of students identified for videotaping after SCDE adjustments (resulting from deletions and replacements) was 232. Videotaping materials were received for 224 of the students. Of these, eight sets of the videotapes were damaged or were excluded from the analysis for other reasons. Additionally, two videotape records could not be linked to operational data. The final number of students in the attained sample was 214. This sample is summarized in Exhibit 38.

Exhibit 38: Demographic Frequencies for the Video Rater Data Sample—by Test Form

	Elementary		Middle		High	
	N	%	N	%	N	%
STUDENT'S ETHNICITY						
African American	52	54.74	40	51.95	23	54.76
American Indian	.	.	.	.	.	.
Asian	2	2.11	2	2.6	.	0
Hawaiian/Pacific Islander	.	.	.	.	.	.
Hispanic	2	2.11	4	5.19	.	0
White	35	36.84	30	38.96	19	45.24
White/African American	1	1.05	1	1.3	.	0
White/American Indian	.	.	.	.	.	.
White/Asian	2	2.11	.	0	.	0
Other	1	1.05	.	0	.	0
STUDENT'S GENDER						
Female	22	23.16	21	27.27	16	38.1
Male	73	76.84	56	72.73	26	61.9
ESL (LANGUAGE)						
Advanced	.	.	.	.	.	.
Beginner	.	.	.	.	.	.
Full English proficient	.	.	.	.	.	.
Intermediate	.	0	1	1.3	.	0
Pre-functional Waiver	2	2.11	5	6.49	.	0
Beginner Waiver	.	.	.	.	.	.
English speaker I	.	0	1	1.3	.	0

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
English speaker II	93	97.89	70	90.91	42	100
Pre-functional	.	.	.	.	.	.
Title III exited	.	.	.	.	.	.
ELIGIBLE FOR FREE OR REDUCED-PRICE LUNCH						
Free	59	62.11	46	59.74	23	54.76
No	29	30.53	27	35.06	18	42.86
Reduced	7	7.37	4	5.19	1	2.38
EFA GRADE (REPORTED GRADE FOR FUNDING)						
1	1	1.05	.	0	.	0
2	5	5.26	.	0	.	0
3	40	42.11	.	0	.	0
4	33	34.74	1	1.3	.	0
5	15	15.79	13	16.88	1	2.38
6	1	1.05	23	29.87	.	0
7	.	0	25	32.47	2	4.76
8	.	0	11	14.29	6	14.29
9	.	0	4	5.19	19	45.24
10	.	0	.	0	13	30.95
11	.	0	.	0	1	2.38
12	.	.	.	.	.	.
COMPLETION STATUS: Student satisfied attemptedness rule						
ELA	95	100	75	97.4	42	100
Mathematics	95	100	75	97.4	42	100
Science[6]	68	71.58	51	66.23	42	100
Social Studies	67	70.53	60	77.92	.	0
COMPLETION STATUS: Student did not answer any content area-items						
ELA	.	0	1	1.3	.	0
Mathematics	.	0	1	1.3	.	0
Science	26	27.37	25	32.47	.	0
Social Studies	28	29.47	16	20.78	42	100
Migrant Status	.	.	.	.	.	.
Home schooled	.	.	.	.	.	.
Medical Homebound	.	0	2	2.6	.	0
IEP DISABILITY CODES (Multiple codes per student)						
Severe Mental Disability	10	10.53	5	6.49	2	4.76
Moderate Mental Disability	29	30.53	30	38.96	23	54.76
Mild Mental Disability	16	16.84	22	28.57	6	14.29
Autism	26	27.37	9	11.69	6	14.29
Deaf/Blindness	0	0	0	0	0	0
Emotional Disability	1	1.05	1	1.3	.	0
Hearing Impairment	7	7.37	4	5.19	2	4.76
Learning Disability	1	1.05	1	1.3	.	0
Multiple Disabilities	4	4.21	4	5.19	6	14.29

[6] The completion rates for science and social studies for the elementary and middle school forms were lower due to sampling of participation in these content areas for two grade-level groups for each form (i.e., students were administered either science or social at these grade levels).

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
Other Health Impairment	6	6.32	1	1.3	5	11.9
Orthopedic Impairment	4	4.21	5	6.49	2	4.76
Speech Language Impairment	78	82.11	42	54.55	19	45.24
Traumatic Brain Injury	.	0	2	2.6	.	0
Visual Impairment	5	5.26	1	1.3	1	2.38
TOTAL	95	100	77	100	42	100

Comparing the attained video rater (VR) sample to the identified sample (see Exhibit 37), the following statements can be made:

By Form

- The attained sample approximates the expected number of students for each form: Elementary students make up 44.4% of the sample, middle school students make up 36.0% of the sample, and high school students make up 19.6% of the sample.

By IEP Disability Code

The first four rows show the primary disabilities of severe, moderate and mild mental disability and autism. These four disabilities are exclusive of each other. In particular, if any of the mental disabilities was coded together with autism, then only the mental disability is reported. Subsequent rows show additional disabilities coded by the test administrators. Since multiple disability codes per student are permitted, their tabled percentages do not add up to 100.

- Severe Mental Disability was sampled similarly to the expectation across forms (elementary school: 10.53%, middle school: 6.49%, and high school: 4.76%).
- Moderate Mental Disability was sampled at rates similar to the expectation in elementary school (30.53%) and middle school (38.96%) and at a higher rate in high school (54.76%).
- Mild Mental Disability was sampled at rates (16.84%, 28.57%, and 14.29%) similar to the expectation.
- Autism was sampled at rates (27.37%, 11.69%, and 14.29%) similar to the expectation.
- The total percentages of students in primary disability categories other than Severe, Moderate, and Mild Disability and Autism were represented at lower rates (14.74%, 14.29%, and 11.90%) than in the identified sample.

Comparing the attained VR sample with the assessed population (see Exhibit 5), the following statements can be made:

By Other Demographic Variables

- For other demographic variables, the proportions in the attained VR sample generally appear to correspond to those seen in the total assessed population when data were available. In the sample, African American (52%-55%),[7] Asian (0%–3%), Hispanic (0%–5%), White (37%–45%), White/African American (0%–1%), White/Asian (0–

[7] The percentage range is reported across the three levels for which there are test forms—elementary, middle, and high school.

2%), and Other (0%–1%) ethnicities were reported. These groups represent the majority of ethnicities in the total population.

- Gender is distributed as approximately two to three males for each female; this ratio is greatest for the elementary school form.
- "English Speaker II" (91% to 100%) reflects the majority of students in the total population.
- Between 55% and 63% of students in the sample were eligible for Free Lunch, approximately the same as in the total population. Slightly fewer students in the sample were also eligible for Reduced Lunch when compared with the total population.
- None of the students in the attained VR sample were home schooled or migrant, and only two students were medical homebound; these results are comparable to the population, which reported rates of typically less than 1% for each of these demographic variables and never more than 2%.

The attained VR sample (Exhibit 38) appears to reasonably represent the identified sample (Exhibit 37) as well as the full population (Exhibit 5). The demographic variables of interest are present in the attained sample data within acceptable ranges of the identified sample and the assessed population.

Item Agreement Analysis

Within each grade band, the absolute difference between test administrator (TA) scores and AIR video rater (VR) scores for each item was computed. Scores that do not differ between TA and VR are noted as "equal"; scores differing by +/–1 score point are noted as "adjacent"; scores differing by more than +/–1 point are flagged as "discrepant." The agreement data are summarized by content area and grade band in Exhibit 39, where values indicate the average percentage of items falling within each agreement category for which there were valid matched responses across TAs and VRs.

Across content areas for the elementary school form, the majority of items (93% to 95%) were shown to be scored as "equal" between the TA and VR; "adjacent" ratings were the next most prevalent (4% to 6%); and "discrepant" ratings were the least prevalent for all content area areas (1%). On the middle school form, all content area areas show a pattern similar to the elementary form. "Equal" categorizations account for the majority of ratings (93% to 94%); the "adjacent" category is next most prevalent (5%); and "discrepant" results account for the smallest proportion of ratings (0% to 2%). Across content areas on the high school form, "equal" ratings again account for the largest proportion (93% to 94%), "adjacent" is the next most prevalent (5% to 6%), then "discrepant" (1% to 2%).

Exhibit 39: Average Item Agreement Statistics by Grade Band and Subject

		Agreement			
Grade-Band	**Subject**	**Equal**	**Adjacent**	**Discrepant**	**Unmatched**
Elementary	ELA	93.04%	5.70%	1.27%	0.00%
	Math	93.13%	5.83%	1.03%	0.00%
	Science	92.56%	6.13%	1.31%	0.00%
	Social Studies	95.00%	3.70%	1.30%	0.00%
Middle	ELA	94.37%	5.03%	0.60%	0.00%
	Math	93.28%	5.15%	1.57%	0.00%
	Science	93.54%	4.91%	1.55%	0.00%
	Social Studies	93.14%	4.83%	2.03%	0.00%
High	ELA	93.13%	5.66%	1.21%	0.00%
	Math	93.62%	4.83%	1.55%	0.00%
	Science	94.38%	4.53%	1.09%	0.00%

Classification Consistency Analysis (as distinct from scoring consistency just discussed in the previous section)

The reported performance levels for each student are derived from a scale score to performance level conversion process. Scale scores are produced based on conversions from the raw scores assigned by the TA. From these scale scores, students were assigned to one of four performance levels (i.e., Levels 1, 2, 3, or 4) within each grade band and content area assessment. Using the VR item scores, correspondence between reported (TA) performance levels and VR performance levels was assessed according to the kappa and weighted kappa coefficients. In ELA, mathematics, science, and social studies, consistency is assessed through *weighted kappa* (Agresti, 1990; Spitzer, Cohen, Fleiss, & Endicott, 1967), which is appropriate for ordered categories:

$$\kappa_w = \frac{\sum\sum w_{ij}\pi_{ij} - \sum\sum w_{ij}\pi_{i+}\pi_{+j}}{1-\sum\sum w_{ij}\pi_{i+}\pi_{+j}},$$

where *i* is the category assigned by the TA, *j* is the category assigned by the VR, $w_{ij} = 1-(i-j)^2/(I-1)^2$ are the weights, π_{ij} is the probability of being classified as *ij*, and "+" indicates agreement between categories. Kappa equals 0 when the agreement is that expected by chance; and kappa equals 1 when there is perfect agreement among raters.

Under the current condition, it must be noted that not all cases included in this analysis contained complete data. Exhibit 40 indicates the *effective sample size* ("n"; cases with information used in the content area-by-form calculation) as well as the *missing count* ("n missing"; indicating students assigned to the current test form with no data for the specified content area).

Exhibit 40: Agreement Statistics by Subject and Grade Band

Grade Band	Subject	κw[8]	95%CI	n / n missing
Elementary	ELA	0.880	0.815 - 0.944	87 / 8
	Mathematics	0.872	0.801 - 0.943	83 / 12
	Science	0.794	0.681 - 0.907	58 / 37
	Social Studies	0.827	0.708 - 0.947	57 / 38
Middle	ELA	0.924	0.864 - 0.983	69 / 8
	Mathematics	0.924	0.864 - 0.984	67 / 10
	Science	0.931	0.867 - 0.995	47 / 30
	Social Studies	0.866	0.754 - 0.977	50 / 27
High	ELA	0.891	0.804 - 0.979	38 / 4
	Mathematics	0.950	0.884 – 1.000	38 / 4
	Science	0.919	0.841 - 0.996	39 / 3
	Social Studies	x	x	x

Summary

TA and VR assignments of students to performance levels typically show high levels of agreement, as weighted kappa typically ranges from 0.79 to 0.95. Further, the 95% confidence intervals show that, while sample sizes for the current calculations may be small, the agreement indices are significantly greater than chance agreement and often approach 1.00. Based on the current evidence, we can conclude that the SC-Alt was accurately scored.

[8] All values are significant at $p < 0.05$.

Appendix D: Descriptions of Achievement Levels (DALs)

English Language Arts Descriptions of Achievement Levels

Performance Level	ELA Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in reading, writing, and communication.	Students performing at level 1 should be able to • attend to a variety of text read aloud as evidenced by facial expressions, gestures, or sounds; • attend to a writing activity using objects, pictures, or letters; • respond to conversations using facial expressions, gestures, or sounds; • attend to a speaker.	Students performing at level 1 should be able to • attend and respond to a variety of text read aloud as evidenced by facial expressions, gestures, or sounds; • demonstrate involvement in a writing activity using objects, pictures, or letters; • participate in conversations as evidenced by facial expressions, gestures, or sounds; • attend and listen to a speaker.	Students performing at level 1 should be able to • respond to a variety of texts read aloud as evidenced by facial expressions, gestures, or sounds; • demonstrate involvement in a writing activity using objects, pictures, or letters; • participate in conversations as evidenced by facial expressions, gestures, or sounds; • attend, listen, and respond to a speaker.
2	Students performing at level 2 demonstrate foundational academic skills and competencies in reading, writing, and communication.	Students performing at level 2 should be able to • participate in reading activities by telling or showing what the text is about, using objects, pictures, or words; • identify individual words; • identify story elements (e.g., main idea, events, setting, and characters); • use oral and written language to describe; • choose topics and generate ideas for written communication; • focus attention on a speaker and listen without interrupting; • participate in conversations by responding appropriately.	Students performing at level 2 should be able to • participate in reading activities by telling or showing what the text is about, using objects, pictures, or words; • participate in reading a variety of texts (e.g., recipes or advertisements); • identify story elements (e.g., main idea, events, setting, characters, and conflict); • make connections within and between texts; • use oral and written language to explain; • choose topics and generate ideas for written communication; • focus attention on a speaker and listen without interrupting; • participate in conversations by responding appropriately; • follow oral and/or written directions.	Students performing at level 2 should be able to • participate in reading activities by telling or showing what the text is about; • participate in reading a variety of texts (e.g., recipes, advertisements, schedules, and newspapers); • identify story elements (e.g., main idea, events, setting, characters, conflict, and plot); • gather meaning from graphic representations; • use oral and written language to explain, inform, and describe; • generate ideas for written communication; • edit own writing; • focus attention on a speaker and listen without interrupting; • participate in conversations by responding appropriately.

English Language Arts Descriptions of Achievement Levels

Performance Level	ELA Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
3	Students performing at level 3 demonstrate increasing academic skills and competencies in reading, writing, and communication.	Students performing at level 3 should be able to • identify story elements in text (e.g., characters, settings, events, cause and effect, and problem solution); • read words and simple sentences; • generate an idea and use words, pictures, or oral language to write; • follow one-step oral or signed directions; • communicate agreement or disagreement appropriately.	Students performing at level 3 should be able to • identify and recall details in text including main idea, plot, characters, and setting; • make predictions about events in text; • determine meaning of unfamiliar words; • generate an idea and use words, pictures, or oral language to write; • follow directions; • initiate conversation.	Students performing at level 3 should be able to • respond to or make connections with text (plot, characters, setting); • make inferences about events in text; • understand multiple meanings of words; • compare and contrast story elements from different stories; • discriminate fact from fiction; • generate an idea and use words, pictures, or oral language to write; • follow directions; • initiate conversation.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in reading, writing, and communication.	Students performing at level 4 should be able to • identify story elements such as the main idea and cause and effect; • make predictions and draw conclusions about text; • read and understand the main idea of a simple paragraph; • create and edit personal written products; • follow multistep oral or signed directions; • take turns appropriately during conversation or discussion.	Students performing at level 4 should be able to • recognize and recall details in text, including the main idea, plot, characters, and setting; • draw conclusions and make predictions and inferences about the text; • read and understand the main idea of a simple paragraph; • explain word meanings; • create and edit personal written products; • follow oral/signed or written directions; • initiate and retell conversations.	Students performing at level 4 should be able to • recognize and recall details in text, including the main idea, plot, characters, and setting; • draw conclusions, and make predictions and inferences about the text; • read and understand the main idea of a short story; • use context clues to understand the meaning of unknown words; • make connections within and between texts and to prior knowledge, other texts, and the world; • create and edit personal written products; • use graphic representations as sources of information.

Mathematics Descriptions of Achievement Levels

Performance Level	Mathematics Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in mathematics.	Students performing at level 1 should be able to • attend to/manipulate one concrete object; • observe that two geometric figures have the same attributes; • observe attributes of objects, such as length and weight.	Students performing at level 1 should be able to • recognize the concept of one in counting objects; • recognize that two geometric figures have the same attributes; • observe attributes of objects, such as length weight and size/volume.	Students performing at level 1 should be able to • recognize the concept of one more in counting objects; • match geometric figures that have the same attributes; • respond to positional concepts such as on top of or under, off-on, above and below; • match objects by one attribute such as length, weight, and size/volume.
2	Students performing at level 2 demonstrate foundational academic skills and competencies in mathematics.	Students performing at level 2 should be able to • count objects in a set; • identify objects by one attribute (color, size, shape); • classify two - and three-dimensional concrete objects according to one attribute; • recognize positional concepts (on/off); • identify measurement tools, including graphs.	Students performing at level 2 should be able to • add and subtract using concrete objects; • classify objects by one attribute (color, size, shape); • recognize and demonstrate understanding of positional concepts (on/off, below/above); • use nonstandard units to measure; • match the correct tool to a specific task (i.e. measure length, weight, time); • identify parts of a chart, graph, or table.	Students performing at level 2 should be able to • solve addition and subtraction problems; • Identify operations (+ or -); • tell which has more in a set; • identify a repeating relationship (pattern); • sort and classify objects by one attribute, (length, height, weight volume); • use a graph or chart to gain information.

Mathematics Descriptions of Achievement Levels

3	Students performing at level 3 demonstrate increasing academic skills and competencies in mathematics.	Students performing at level 3 should be able to • demonstrate addition and subtraction concretely or symbolically; • count and compare objects in a set; • sort and classify objects by attribute (shape, size); • identify three-dimensional shapes (cube, sphere, cylinder); • use nonstandard units to measure; • find answers to questions in a graph.	Students performing at level 3 should be able to • identify the answer to one-digit addition and subtraction problems; • identify a set as having more, fewer, or the same number as another set; • identify and extend a repeating pattern; • compare three-dimensional shapes by attribute; • compare length of two objects (shorter/longer); • interpret information displayed in a graph.	Students performing at level 3 should be able to • identify the process for solving an addition or a subtraction problem; • identify and use operational symbols correctly; • estimate the number of objects in a set; • add to find value of a set of coins; • describe, create, and complete a repeating pattern; • use and organize data to create charts, graphs, and tables.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in mathematics.	Students performing at level 4 should be able to • demonstrate understanding of addition and subtraction; • generate a pattern using three-dimensional shapes (cube, sphere, cylinder); • compare objects by attribute (length, size); • interpret information displayed in a graph.	Students performing at level 4 should be able to • solve addition and subtraction facts without regrouping; • identify, describe, and extend a repeating pattern; • interpret information displayed in a graph; • use data to create graphs or tables.	Students performing at level 4 should be able to • identify, compare, and construct numbers; • use operation symbols (more than less than and equal to) to solve problems; • add to find the value of a set of two or more coins; • identify, describe, create, extend, and complete a repeating pattern; • describe events as more likely or less likely to occur; • use and organize data to create and interpret graphs.

Science Descriptions of Achievement Levels

Performance Level	Science Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in science.	Students performing at level 1 should be able to • attend to a science investigation; • observe sequence of growth (e.g., young and old); • attend to daily weather conditions; • recognize sun and moon in reference to day and night; • observe objects in motion.	Students performing at level 1 should be able to • attend and participate in a scientific investigation; • identify major body parts of animals; • identify sun and moon; • observe the motion of objects; • sort by one attribute.	Students performing at level 1 should be able to • attend and respond to a scientific investigation; • attend to objects moved by force; • observe that an object at rest moves.
2	Students performing at level 2 demonstrate foundational academic skills and competencies in science.	Students performing at level 2 should be able to • participate in a scientific investigation; • distinguish young from old; • identify daily weather conditions; • match appropriate activities to day and night (go to school during the day/sleep at night); • identify the position of objects such as above/below, inside, or on top; • describe materials by observable properties.	Students performing at level 2 should be able to • predict the results of a scientific investigation; • sort and describe materials by observable properties; • match major organs of animals to their function; • identify the pattern of day and night; • identify if an object is moving; • identify the role of a switch in a simple electrical circuit.	Students performing at level 2 should be able to • explain information or events based on observation; • identify the force that makes an object move; • predict the outcome of a scientific investigation related to electricity or force and motion.
3	Students performing at level 3 demonstrate increasing academic skills and competencies in science.	Students performing at level 3 should be able to • classify events in sequential order; • conduct a simple scientific investigation; • match a tool to the task; • identify living and nonliving things; • identify major organs of animals; • compare daily changes in weather conditions; • identify water in solid and liquid form; • identify the temperature on a thermometer as hot or cold.	Students performing at level 3 should be able to • predict the outcome of a scientific investigation and compare the results with the prediction; • read data from simple tools; • use graphs, tables, or diagrams to gain information; • identify the characteristics of living and nonliving things; • identify what plants need to grow; • identify functions of major organs of animals; • identify the changes in the seasons.	Students performing at level 3 should be able to • predict the outcome of a scientific investigation and compare the results with the prediction as they relate to force and motion, friction and gravity; • compare magnetic and nonmagnetic objects; • identify electricity as a source of energy; • relate the change in force to the change in speed.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in science.	Students performing at level 4 should • gain meaning from graphs and tables; • conduct and analyze the results of a scientific investigation; • identify major organs of animals and their functions; • identify living and nonliving things in terms of a food web; • identify natural resources as renewable or nonrenewable; • identify how heat and light change from season to season.	Students performing at level 4 should be able to • conduct and analyze the results of a scientific investigation; • gain meaning from graphs, tables, or diagrams; • describe what plants need to survive; • describe temperature ranges; • identify simple machines (inclined plane, lever, pulley); • identify how heat and light change from season to season.	Students performing at level 4 should be able to • plan, conduct, and analyze the results of a scientific investigation; • identify how simple machines are used to help people (inclined plane, lever, pulley, etc.); • predict and identify the effect of the change in force on an object; • describe water as solid, steam, or liquid; • investigate how to increase the speed of a falling object.

Social Studies Descriptions of Achievement Levels

Performance Level	Social Studies Achievement Level Definitions	Grades 3–5	Grades 6–8
1	Students performing at level 1 demonstrate emerging academic skills and competencies in social studies.	Students performing at level 1 should be able to • identify self from others • respond to a person in authority in the home or school; • follow class rules; • engage in turn-taking; • attend to information presented orally about South Carolina history.	Students performing at level 1 should be able to • identify self from others; • respond to familiar authority figures; • follow class rules; • engage in turn-taking and sharing; • respond to information about significant and historical events in South Carolina.
2	Students performing at level 2 demonstrate foundational skills and competencies in social studies.	Students performing at level 2 should be able to • identify characteristics such as gender that help identify self in relation to others; • match workers to different jobs in the community; • recognize people in authority and follow class rules; • match the people we honor on some national holidays (e.g., George Washington, Martin Luther King, Jr.) with the holidays; • match jobs of the past with jobs of the present; • match significant historical figures such as Thomas Edison to their inventions.	Students performing at level 2 should be able to • identify surroundings (e.g., classroom, school); • match different people to their jobs in the community; • identify people in authority and follow class rules; • demonstrate understanding of rules; • identify the people we honor on some national holidays (e.g., George Washington, Martin Luther King, Jr.); • identify the purpose of money; • match changes over time to the past and present such as communication.
3	Students performing at level 3 demonstrate increasing skills and competencies in social studies.	Students performing at level 3 should be able to • understand the concept of past and present; • demonstrate respect for people in authority; • identify major symbols of the United States; • identify why we celebrate the national holidays; • recognize that when we work we earn money to buy things; • identify features on a map of South Carolina (river, mountain, ocean); • answer questions about significant events related to the Civil War; • match accomplishments to historical figures such as Thomas Edison, Alexander Graham Bell, etc.	Students performing at level 3 should be able to • identify members of the larger community (e.g., police officers, fire-fighters, doctors); • demonstrate understanding of consequences of not following the rules; • Identify examples of good citizenship such as honesty, courage, etc.; • identify symbols of the United States (e.g., the flag, bald eagle); • demonstrate an understanding that we work to earn money and use money to buy things; • identify changes over time such as in travel, farming, etc.; • gain information from maps, charts, and graphs; • answer questions about key historical figures and significant historical events including the civil rights movement.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in social studies.	Students performing at level 4 should be able to • place personal history on a time line; • identify the roles of leaders and officials in local government (e.g., principal, mayor, governor); • identify individuals who embody qualities of good citizenship; • identify examples of respect and fair treatment; • recognize that we exchange money for goods and services; • use a key to locate geographic features on a map of South Carolina; • answer questions about key concepts related to the Civil War; • answer questions about the accomplishments of key historical figures such as Thomas Edison, Alexander Graham Bell, etc.	Students performing at level 4 should be able to • place personal and family history on a time line; • Identify roles of leaders and officials in local government (e.g., principal, mayor, governor) • identify examples of the qualities of courage and patriotism; • identify examples of respect and fair treatment and their opposites; • recognize how the amount of money available determines what we can buy; • gain information from maps and charts; • identify the accomplishments of Civil Rights leaders including Rosa Parks.

Appendix E: Summary of Linking Design

How South Carolina Alternate Assessment Standards and Measurement Guidelines (ASMGs) Overlap across Grade Bands

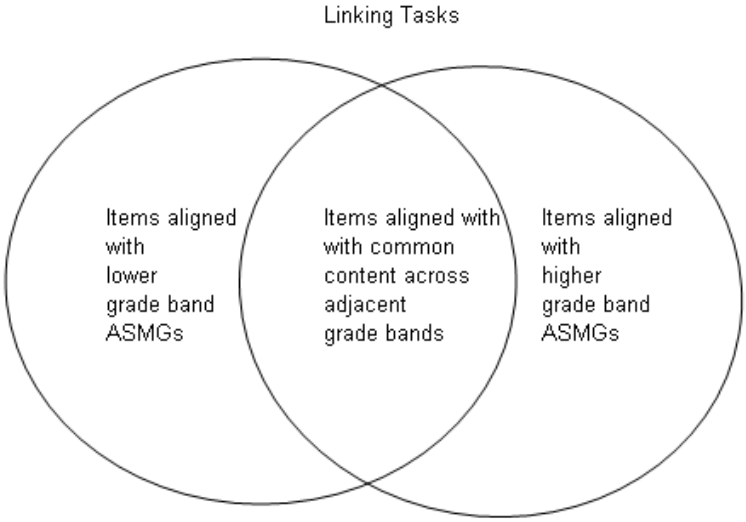
Common threads across grade bands were targeted in the development of some ASMGs to promote consistent instruction across the curriculum from grade band to grade band. The difference in the essence of some standards is subtle in a number of academic standards across grade bands, and in fact some standards are exactly the same in some of the content standards.

Structure of the tasks

- Each task has four to six items. The student responses to each item are scored from one to four points depending on demands of the response.
- Only one beginning item in the first three to five tasks in an academic content area is an engagement item. Each of the engagement items is aligned with the academic content standards through the ASMGs. The remaining items of the tasks are aligned to the academic content standards through the ASMGs at complexity levels ranging from low to high. Since every student must respond to all items in at least a minimum of five tasks, every student must respond to items that assess his or her knowledge of content and skills at the grade band to which he or she is assigned.
- Items and tasks progress upward in complexity and difficulty across the performance levels at the assigned grade band.

Structure of linking tasks

- ASMGs from adjacent grade bands were examined for common threads linked to content across the two grade bands for use in developing linking tasks.
- Some items were developed specifically to link to ASMGs that were common in academic demand across grade bands. Other items were designed specifically to assess only the ASMG content for a specific grade band.



Subject	Grade Band	Number of Items	Number of Tasks	**Starting Positions**									
				Starting Task 1		Within Grade-Band Linking		Starting Task 3		Within Grade-Band Linking		Starting Task 6	
				Items	Tasks	Items	Tasks	Items	Tasks	Items	Tasks	Items	Tasks
ELA	3–5	68	12	32	5	19	3	38	7	19	4	36	7
	(Linking)	*29*											
	6–8	65	12	31	5	17	3	39	7	22	4	34	7
	(Linking)	*24*											
	10	64	12	28	5	16	3	37	7	21	4	36	7
Mathematics	3–5	53	12	23	5	14	3	30	7	16	4	30	7
	(Linking)	*37*											
	6–8	55	12	23	5	14	3	31	7	17	4	32	7
	(Linking)	*29*											
	10	60	12	24	5	15	3	34	7	19	4	36	7
Science	3–5	58	12	26	5	16	3	35	7	19	4	32	7
	(Linking)	*25*											
	6–8	60	12	28	5	17	3	36	7	19	4	32	7
	(Linking)	*20*											
	10	56	12	26	5	15	3	33	7	18	4	30	7
Social Studies	3–5	57	13	24	6	16	4	32	8	20	5	37	8
	(Linking)	*24*											
	6–8	55	13	24	6	16	4	33	8	21	5	35	8

Appendix F: Item Statistics Summaries for the Spring 2008 Field-Test Social Studies Items

Grade Band 3–5 Social Studies Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1093	54	0.49	0.38	0.00	1.82	+A	+A
1092	55	0.48	0.43	1.45	1.62	-A	-A
1094	56	0.45	0.40	0.73	1.46	-A	-A
1091	57	0.57	0.40	0.00	0.71	+B	+A

Grade Band 6–8 Social Studies Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1093	52	0.47	0.45	0.00	2.75	+A	+A
1092	53	0.49	0.53	0.00	1.04	-A	+A
1094	54	0.42	0.46	0.00	2.10	-A	+A
1091	55	0.53	0.49	0.00	2.59	-A	+A

Social Studies: Field-Test WINSTEPS Item Statistics

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
54	0.90	1222	1244	0.04	1.22	6.20	1.31	5.66	ITS_item_1093
55	0.48	1222	1448	0.05	1.16	4.51	1.14	2.92	ITS_item_1092
56	0.78	1222	1290	0.05	1.33	8.66	1.32	7.16	ITS_item_1094
57	0.75	1222	1339	0.04	0.93	-2.17	0.88	-1.82	ITS_item_1091

Appendix G: Marginal Reliability by Grade Band, Subject, and Starting Task

Table G-1. Marginal Reliability by Starting Task and Grade Band for ELA

Grade Band	Initial Task	N	Reliability	$\overline{\sigma}_{e^*}$
Grades 3-5	1	405	0.84352	26.42034
	3	290	0.87676	10.91223
	6	517	0.76893	24.87322
Grades 6-8	1	299	0.87287	24.31598
	3	184	0.90221	11.42836
	6	560	0.74746	27.45807
Grade 10	1	109	0.87653	29.21860
	3	63	0.88247	15.18382
	6	200	0.79281	27.92301

Table G-2. Marginal Reliability by Starting Task and Grade Band for Mathematics

Grade Band	Initial Task	N	Reliability	$\overline{\sigma}_{e^*}$
Grades 3-5	1	386	0.79223	27.72609
	3	345	0.84742	13.14552
	6	485	0.76431	24.69769
Grades 6-8	1	293	0.81304	26.60262
	3	217	0.84859	12.74215
	6	529	0.75632	28.26488
Grade 10	1	108	0.83471	31.84190
	3	81	0.88379	11.46787
	6	177	0.80438	23.92433

Table G-3. Marginal Reliability by Starting Task and Grade Band for Science

Grade Band	Initial Task	N	Reliability	$\overline{\sigma}_{e^*}$
Grades 3-5	1	350	0.82290	26.19885
	3	267	0.84738	14.04201
	6	250	0.75166	19.14248
Grades 6-8	1	259	0.85479	26.69059
	3	143	0.87183	15.86810
	6	354	0.72476	25.18742
Grade 10	1	131	0.82575	35.93906
	3	56	0.84204	19.98707
	6	179	0.76022	28.43984

Table G-3. Marginal Reliability by Starting Task and Grade Band for Social Studies

Grade Band	Initial Task	N	Reliability	$\overline{\sigma}_{e^*}$
Grades 3-5	1	254	0.76789	35.02361
	3	132	0.82481	13.98038
	6	310	0.78857	21.49156
Grades 6-8	1	192	0.81995	32.93442
	3	84	0.86141	14.09208
	6	338	0.70909	30.35977

Appendix H: Score Report Sample

INDIVIDUAL STUDENT REPORT

Prepared Especially for the Family of
Kyree Adams

Date of Birth:	09-20-1997
Student ID:	123456123456
School District:	Columbia
School:	Alfonso Elementary School

Spring 2008

The South Carolina Alternate Assessment (SC-Alt)

Kyree participated in the South Carolina Alternate Assessment (SC-Alt) during the spring of 2008.

She took the elementary form of the test, which is based on academic standards from grades 3 to 5. This report is designed to provide you with information on your child's performance on this assessment.

The SC-Alt is a test designed for students with significant cognitive disabilities who participate in a school curriculum that includes academic and functional skill instruction. The alternate assessment only tests students' achievement in English language arts (ELA), mathematics, science, and social studies. Individualized Education Program (IEP) reports and other methods provide parents with information on how students are progressing in the other areas.

What is the SC-Alt?

- The SC-Alt assessment includes performance tasks in each subject area. Students may complete the tasks by using their usual method of communication. This may include pointing or gazing at answer choices, selecting objects, pictures, or picture symbols that represent an answer choice, or reading letters, words or sentences to complete the task.
- The tasks are linked to the state academic content standards in four areas: English language arts (ELA), mathematics, science, and social studies.
- Students are assigned a test form based on their age. Students ages 8-10 are assigned to the elementary school form; students ages 11-13 are assigned to the middle school form; and students age 15 take the high school form.

How are scores reported and used?

- Four achievement levels (Level 1, Level 2, Level 3, and Level 4) have been established for the SC-Alt. Achievement levels describe how students are doing in relation to the state academic standards. Your child's performance is also reported as a scale score that allows parents to monitor growth from year to year.
- Level 2 is the achievement level reported as meeting the "Basic" reporting requirement for state accountability on the District Report Card. Achievement Levels 3 and 4 are the achievement levels reported as "proficient" for schools and districts in the federal accountability Adequate Yearly Progress (AYP) report.

Where can I get more information about SC-Alt and my child's performance?

- You can contact your child's teacher or school for more information.
- You can view examples of tasks, information about expectations at each achievement level, and scale score tables on the South Carolina Department of Education website at http://www.ed.sc.gov/agency/offices/assessment/programs/SWD/SouthCarolinaAlternateAssessmentSC-Alt.html.

The South Carolina Department of Education

Spring 2008
Kyree Adams

The SC-Alt

Mathematics

Kyree scored at Level 3 with a scale score of 480 in mathmatics.

Students who score at Level 3 should be able to:

- add and subtract simple numbers;
- count and compare objects in a group;
- compare objects by color, size, or shape;
- identify three-dimensional shapes;
- read information in a graph.

	4	Students performing at Level **4** demonstrate and apply academic skills and competencies in mathematics.
Your Child's Level ▶	3	Students performing at Level **3** demonstrate increasing academic skills and competencies in mathematics.
	2	Students performing at Level **2** demonstrate foundational academic skills and competencies in mathematics.
	1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in mathematics.

How you can support Kyree's learning

- Encourage your child to add and subtract during everyday activities. For example, show her 5 cookies or other objects and take 2 away. Then, ask her to tell you how many are left (5-2=3).
- Play games that require counting spaces and adding points for the score. For example, if you spin a 5, your
- Describe everyday household objects by shapes. For example, the window is a large rectangle.

English Language Arts

Kyree scored at Level 4 with a scale score of 495 in ELA.

Students who score at Level 4 should be able to:

- identify the main idea and make predictions about what will happen next in a story;
- write a simple story;
- follow multi-step directions;
- take turns appropriately during conversations.

Your Child's Level ▶	4	Students performing at Level **4** demonstrate and apply academic skills and competencies in reading, writing and communication.
	3	Students performing at Level **3** demonstrate increasing academic skills and competencies in reading, writing and communication.
	2	Students performing at Level **2** demonstrate foundational academic skills and competencies in reading, writing and communication.
	1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in reading, writing and communication.

How you can support Kyree's learning

- Encourage your child to read passages from a variety of materials (books, magazines, or newspapers).
- Read a story with your child and talk to her about specific characters and the order of events in the story. Ask her what she thinks will happen next.
- Assist your child in writing a letter or note to a family member or friend. You can use pictures, words, or tactile means, such as Braille.
- Encourage your child to participate in conversations with friends and family using words, pictures, or assistive technology devices.

2

Spring 2008
Kyree Adams

Science

Kyree scored at Level 1 with a scale score of 270 in science.

Students who score at Level 1 should be able to:
- observe daily weather conditions;
- recognize day and night;
- observe objects in motion.

4	Students performing at Level 4 demonstrate and apply academic skills and competencies in science.
3	Students performing at Level 3 demonstrate increasing academic skills and competencies in science.
2	Students performing at Level 2 demonstrate foundational academic skills and competencies in science.
Your Child's Level ▶ 1	Students performing at Level 1 may demonstrate emerging academic skills and competencies in science.

How you can support Kyree's learning

- Talk to your child about the weather conditions each day. For example, take her outside to feel the rain.
- Point out the moon in the sky.
- Talk to your child about what we do in the daytime and at night. For example, we sleep at night.
- Roll a ball across a table or the floor and talk about how fast it is going.

Social Studies

Kyree scored at Level 1 with a scale score of 270 in social studies.

Students who score at Level 1 should be able to:
- identify self as separate from others;
- respond to familiar authority figures (home, school, and community);
- follow rules;
- engage in turn-taking and sharing;
- begin to distinguish between the past and the present.

4	Students performing at Level 4 demonstrate and apply academic skills and competencies in social studies.
3	Students performing at Level 3 demonstrate increasing academic skills and competencies in social studies.
2	Students performing at Level 2 demonstrate foundational academic skills and competencies in social studies.
Your Child's Level ▶ 1	Students performing at Level 1 may demonstrate emerging academic skills and competencies in social studies.

How you can support Kyree's learning

- Through a variety of activities have your child respond to her name.
- Have your child follow a simple direction given by the person in authority.
- Put a simple chart on the refrigerator and add stars or stickers to reward following simple rules.
- Help your child take turns in a simple game or family activity.
- Talk about things that happened in the past and things that are happening today (e.g., yesterday we ate chicken for supper—today we are eating hamburgers; yesterday we went to the grocery store—today we are going to the park).

3

Spring 2008
Kyree Adams

The SC-Alt

The following areas are tested in Mathematics:

Number and Operations
- whole numbers
- fractions
- addition and subtraction
- multiplication and division

Algebra
- patterns and their relationships

Geometry
- attributes of objects such as shape, size, color
- identification of two- and three-dimensional shapes

Measurement
- money
- length, liquid, volume, and mass and weight
- time
- equivalences

Data Analysis and Probability
- data collection and representation
- data analysis
- probability

The following areas are tested in English Language Arts:

Reading
- reading
- comprehending a variety of texts (such as fiction, nonfiction, poetry, and drama)
  Note: Reading materials may include objects, pictures or photographs, picture symbols, letters, and words.

Writing
- developing written communications (notes, stories) using the student's typical method of communication

Communication
- speaking*
- listening

**Students' typical method of communication, verbal or nonverbal, may be facilitated by using objects, pictures or photographs, picture symbols, letters and words, voice output devices, or assistive technology.*

The following areas are tested in Science:

Scientific Inquiry involves studying scientific processes and skills such as:
- observing
- classifying
- predicting what will happen in a simple scientific experiment

Life Science
- basic needs of plants and animals
- their structures and habitats

Earth Science
- weather
- objects in the sky (sun and moon)
- earth materials (rocks and soil)

Physical Science
- characteristics of objects
- the effect of force on the motion of objects
- light, heat, and electricity

The following areas are tested in Social Studies:

Social Studies Literacy Elements are concepts required for understanding this subject such as:
- distinguishing between past, present, and future
- demonstrating responsible citizenship within the school community, the local community, and national communities
- creating and using timelines
- understanding the relationship between people and the land

Academic Standards include concepts related to specific historical time frames:
- history
- geography
- political science/government
- economics

4